

Université de Sherbrooke

Exploration des déterminants structuraux de la signalisation biaisée du
récepteur de l'apeline

Par
Laurent Bruneau-Cossette
Programme de maîtrise en biochimie

Mémoire présenté à la Faculté de médecine et des sciences de la santé
en vue de l'obtention du grade de maître ès sciences (M. Sc.)
en biochimie

Sherbrooke, Québec, Canada
Septembre 2019

Membres du jury d'évaluation
Éric Marsault, Directeur, Département de pharmacologie et physiologie,
Université de Sherbrooke
Pierre Lavigne, Co-directeur, Département de biochimie, Université de
Sherbrooke
Michelle Scott, Évaluatrice interne au programme, Département de biochimie,
Université de Sherbrooke
Armand Soldera, Évaluateur externe au programme, Département de chimie,
Faculté des sciences, Université de Sherbrooke

© Laurent Bruneau-Cossette, 2019

Résumé

Les récepteurs couplés aux protéines G sont une classe de protéines membranaires de haute importance d'un point de vue pharmacologique. Leurs actions dans l'organisme, allant de l'odorat à la régulation du métabolisme du glucose, passent toutes par un phénomène plus global : la transduction du signal. Ainsi, les récepteurs couplés aux protéines G transforment un signal extracellulaire, sous la forme d'un ligand, en un signal intracellulaire, principalement via l'action des protéines G et des arrestines, générant ultimement une réponse de l'organisme à un stimulus. Cette fonction, et donc leur importance, est réalisée grâce à la nature dynamique de leur structure : les changements conformationnels de ces récepteurs sont essentiels à leur fonction. Il est désormais admis que les récepteurs couplés aux protéines G peuvent provoquer des réponses cellulaires différentes en fonction du ligand auquel ils sont exposés, par exemple en stimulant davantage les protéines G que les arrestines et inversement. Il s'agit du phénomène de la signalisation biaisée. Or, la nature des changements conformationnels de ces récepteurs pouvant mener à une signalisation biaisée est encore mal connue. L'objectif de ce travail est d'étudier les déterminants structuraux associés à la signalisation biaisée en utilisant comme modèle le récepteur de l'apeline, un récepteur impliqué dans la régulation du système cardiovasculaire dont le profil de signalisation est bien connu. En utilisant une approche combinée d'analyse théorique de simulations de dynamique moléculaire et d'analyse expérimentale du profil fonctionnel de mutants du récepteur, les déterminants structuraux de l'affinité du récepteur pour la protéine $G\alpha_{i1}$ sont maintenant connus avec plus de détails. Au cours du processus, un protocole d'analyse et d'échantillonnage de simulations de dynamique moléculaire a été élaboré et testé avec succès. De plus, cinq nouveaux mutants du récepteur dont les profils de signalisation diffèrent substantiellement du profil du récepteur de type sauvage ont été identifiés, dont deux possédant une activité basale (sans stimulation par un ligand) exclusive à la protéine $G\alpha_{i1}$. Bien qu'il manque certains résultats pour exploiter le plein potentiel des résultats de ce travail et que certains aspects du travail peuvent être substantiellement améliorés, les développements présentés ici dressent un portrait clair de la signalisation biaisée du récepteur de l'apeline et des GPCRs en général et ouvrent plusieurs avenues de recherche qui pourront être empruntées dans les années à venir.

Mots clés : GPCR, signalisation biaisée, apeline, APJ, dynamique moléculaire, BRET

Table des matières

Résumé	iii
Liste des figures	x
Liste des tableaux	xii
Liste des abréviations	xiii
Remerciements	xv
Introduction: biochimie des GPCRs	1
Les GPCRs	1
Fonction et intérêt	1
Génétique et diversité	2
La structure des GPCRs	3
Données actuelles	3
Notation Ballesteros-Weinstein	7
Interactions conservées entre les domaines transmembranaires	8
Domaines extracellulaires et intracellulaires	10
La signalisation des GPCRs	12
Chronologie	12
Les protéines G	14
Les RGSs	17
Les GRKs	17
Les arrestines	19
Étude de la fonction des GPCRs	21
Descripteurs recherchés du comportement des GPCRs et de leurs ligands	21
Liaison du ligand au récepteur et essais de liaison par saturation	22
Essais fonctionnels de la signalisation des GPCRs : le modèle opérationnel	24

La signalisation biaisée	29
Signalisation et effets physiologiques : efficacité pluridimensionnelle	29
Le biais systémique	30
Le biais observationnel.....	30
Le biais ligand-dépendant	31
Évaluation du biais systémique associé à une mutation sur un récepteur	33
La transduction du signal par les GPCRs.....	34
Chronologie et enjeux	34
Études préalables de mutagenèse	34
Le Major H-Bond Network et le motif NPxxY	36
Le motif DRY.....	36
Le Hydrophobic Hindering Mechanism et le motif WxP (tryptophane commutateur)	37
Dynamique des régions importantes dans l'activation des GPCRs.....	37
Interaction des GPCRs avec leurs partenaires de liaison	39
Structure et signalisation biaisée	41
Le récepteur de l'apeline.....	42
Profil de signalisation, fonctions principales et relation entre signalisation et effets physiologiques.....	42
Structure d'APJ et relation structure-activité de ses ligands	43
Présentation du projet.....	44
Déterminants structuraux de la signalisation biaisée d'APJ : un cycle de l'ignorance	44
Approche et objectifs.....	46
Matériel et méthodes.....	48
Présentation des méthodes théoriques utilisées	48
La dynamique moléculaire : théorie et pratique	48

Limitations intrinsèques, limitations techniques et comment y remédier	51
L'analyse des simulations de dynamique moléculaire	55
Présentation des méthodes expérimentales utilisées	77
Les essais de liaison par compétition	77
Les ELISAs de surface	80
Les essais BRET	81
Dynamique moléculaire des systèmes hAPJ-apeline et hAPJ	84
Construction des systèmes hAPJ-apeline et hAPJ	84
Dynamique moléculaire de production	85
Analyse des simulations de dynamique moléculaire : TOCATA-MD	86
Simplification des données : tICA/GMHMM	86
Analyse des données : sous-matrices principales	88
Analyse des données : distance de Bhattacharyya	92
Gestion semi-automatique des simulations en série : fonctionnalités de TOCATA-MD	94
Génération des mutants de hAPJ	95
Matériel de mutagenèse par QuickChange, transformation bactérienne et extraction d'ADN	95
Méthode de mutagenèse par QuickChange	96
Méthode de transformation bactérienne et de purification des plasmides	97
Analyse expérimentale des mutants de hAPJ	98
Matériel générique pour la culture cellulaire et les expériences associées	98
Méthode de culture cellulaire de routine	99
Matériel spécifique à l'essai de liaison par compétition	99
Méthode des essais de liaison	100
Matériel spécifique aux ELISAs de surface	101

Méthode des ELISAs de surface	102
Matériel spécifique aux essais BRET.....	103
Méthode des essais BRET	103
Résultats.....	106
Analyse des systèmes hAPJ-apeline et hAPJ-apo.....	106
Analyse théorique du système hAPJ-apeline	106
Analyse théorique du système hAPJ-apo	111
Design de mutants de hAPJ	115
Analyse fonctionnelle des mutants sélectionnés	116
Essais de liaison par compétition	116
ELISAs de surface.....	118
Essais BRET.....	120
Analyse théorique du système hAPJ-D75N-apo.....	124
Analyse théorique du système hAPJ-N112G-apo.....	129
Analyse théorique du système hAPJ-D126V-apo.....	132
Analyse théorique du système hAPJ-V254L-apo	136
Analyse comparative théorique des systèmes mutants	139
<i>Comparaison des systèmes hAPJ-WT-apo et hAPJ-D75N-apo</i>	139
<i>Comparaison des systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ-N112G-apo.</i>	143
<i>Comparaison des systèmes hAPJ-WT-apo et hAPJ-D126V-apo</i>	147
<i>Comparaison des systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ-V254L-apo</i> .	150
Discussion.....	154
Résumé des résultats	154
Critique du protocole d'échantillonnage.....	154
Critique de la tICA comme outil d'analyse des données	156

Critique de l'analyse par sous-matrices principales de la tICA	158
Critique des HMM en phase d'échantillonnage	159
Critique de l'utilisation de la distance de Bhattacharyya pour séparer les états	161
Critique des paramètres structuraux utilisés.....	163
Analyse des résultats expérimentaux	166
Critique de la qualité et limitations des résultats expérimentaux	166
Analyse de la qualité et la validité des résultats des essais de liaison compétitifs	169
Analyse de la qualité et la validité des résultats des essais BRET	173
Relation structure-activité des mutants hAPJ en relation à l'effet du sodium	176
Analyse comparative des systèmes mutants	180
Limitations de l'analyse théorique des différents systèmes de hAPJ.....	180
Comparaison des systèmes hAPJ-WT-apeline et hAPJ-WT-apo.....	184
Analyse comparative des systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ mutants	185
Conclusion	189
Liste des références.....	191
Annexe I : Courbes de déplacement de l'apeline radiomarquée par de l'apeline non- radiomarquée.	205
Annexe II : Courbes dose-réponse de BRET $G\alpha_{i1}$ et β -arrestine 2 pour l'ensemble des mutants étudiés.	217
Annexe III : Inventaire des mutants d'intérêt répertoriés dans la littérature sur les GPCRs	225
Annexe IV : Résultats de l'analyse par sous-matrices principales du système hAPJ-apo	234
Annexe V : Résultats de l'analyse par sous-matrices principales du système hAPJ-D75N- apo.....	239

Annexe VI : Résultats de l'analyse par sous-matrices principales du système hAPJ-N112G- apo.....	247
Annexe VII : Résultats de l'analyse par sous-matrices principales du système hAPJ- D126V-apo	251
Annexe VIII : Résultats de l'analyse par sous-matrices principales du système hAPJ- V254L-apo	257

Liste des figures

Figure 1 : Structure et régions importantes d'un GPCR.....	4
Figure 2 : Localisation des résidus n.50 sur le récepteur hAPJ.....	8
Figure 3 : Contacts non-covalents conservés chez les GPCRs.....	10
Figure 4 : Chronologie de la signalisation des GPCRs.....	12
Figure 5 : Interaction entre une protéine G et un GPCR.....	15
Figure 6 : Interaction entre une arrestine et un GPCR.....	20
Figure 7 : Régions des GPCRs rhodopsin-like impliquées dans la transduction du signal..	35
Figure 8 : Interfaces d'interaction entre les GPCRs et les protéines G et les arrestines.....	40
Figure 9 : Exemple de groupage par <i>k-means</i> sur les pixels d'une image.....	66
Figure 10 : Exemple de groupage par mélange de Gaussiennes	67
Figure 11 : Comparaison d'une chaîne de Markov et d'un HMM	69
Figure 12 : Description des étapes d'un ELISA de surface.....	81
Figure 13 : Fonctionnement du BRET pour les interactions entre un GPCR et les protéines G et les arrestines	83
Figure 14 : Description du protocole de base de TOCATA-MD	95
Figure 15: Résultats du programme d'analyse de base du système hAPJ-WT-apeline	107
Figure 16 : Résultat de l'analyse par sous-matrices principales de la composante indépendante 1 du système hAPJ-WT-apeline	108
Figure 17 : Résultat de l'analyse par sous-matrices principales de la composante indépendante 2 du système hAPJ-WT-apeline	109
Figure 18 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-WT-apeline.....	110
Figure 19 : Résultats du programme d'analyse de base du système hAPJ-WT-apo	113
Figure 20 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-WT-apo	115
Figure 21 : Données expérimentales représentatives d'un essai de liaison par compétition	117
Figure 22 : Expression relative des mutants de 2HA-hAPJ à la surface de la cellule mesurée par ELISA de surface.....	119
Figure 23 : Données brutes représentatives d'un essai BRET.....	123
Figure 24 : Résultats du programme d'analyse de base du système hAPJ-D75N-apo.....	126
Figure 25 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-D75N-apo.....	128
Figure 26 : Résultats du programme d'analyse de base du système hAPJ-N112G-apo....	130
Figure 27 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-N112G-apo.....	131
Figure 28 : Résultats du programme d'analyse de base du système hAPJ-D126V-apo....	134
Figure 29 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-D126V-apo.....	135
Figure 30 : Résultats du programme d'analyse de base du système hAPJ-V254L-apo	137
Figure 31 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-V254L-apo.....	138
Figure 32 : Différences entre deux conformations choisies des systèmes hAPJ-WT-apo et hAPJ-D75N-apo	140

Figure 33 : Distributions d'angles dièdres d'intérêt pour la comparaison des systèmes hAPJ-WT-apo et hAPJ-D75N-apo	142
Figure 34 : Illustration des deux changements conformationnels majeurs observés seulement dans les systèmes hAPJ-WT-apeline et hAPJ-N112G-apo	144
Figure 35 : Distributions d'angles dièdres d'intérêt pour la comparaison des systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ-N112G-apo	146
Figure 36 : Illustration des changements conformationnels distinguant le système hAPJ-D126V-apo du système hAPJ-WT-apo	148
Figure 37 : Distribution des angles dièdres expliquant le changement conformationnel du TM3 pour le système hAPJ-D126V-apo	149
Figure 38 : Comparaison et explication des différences entre les systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ-V254L-apo	151
Figure 39 : Distribution des angles dièdres expliquant le changement conformationnel du TM3 pour le système hAPJ-V254L-apo	153
Figure 40 : Illustration des concepts d' <i>induced fit</i> , de sélection conformationnelle et d'instabilité induite par mutation	171
Figure 41 : Explication de l'effet du sodium sur l'activité du récepteur de l'apeline	179

Liste des tableaux

Tableau 1 : Codes PDB représentatifs de structures publiées sur chaque GPCR dont au moins une structure existe sur la PDB.....	5
Tableau 2 : Nombre de contacts conservés totaux et intracellulaires entre chaque paire de domaines transmembranaires.....	9
Tableau 3 : Proportion de mutations altérant le profil fonctionnel des GPCRs dans les régions susceptibles d'être impliquées dans l'activation des GPCRs.....	35
Tableau 4 : Paramètres d'intérêt pour l'étude du déplacement de l'apeline radiomarquée par l'apeline non-radiomarquée par régression non-coopérative.	116
Tableau 5 : Paramètres d'intérêt pour l'étude de l'expression des mutants à la surface de la cellule par ELISA de surface.	118
Tableau 6 : Paramètres d'intérêt pour l'étude du recrutement de la protéine $G\alpha_{i1}$ par les mutants de 2HA-hAPJ	121
Tableau 7 : Paramètres d'intérêt pour l'étude du recrutement de la protéine β -arrestine 2 par les mutants de 2HA-hAPJ.....	121

Liste des abréviations

AA : Acide aminé
 ADN : Acide désoxyribonucleique
 AMPc : Adénosine monophosphate cyclique
 ANOVA : Analyse de variance
 AP2 : Protéine adaptatrice 2 (*Adaptor Protein 2*)
 BIC : Critère d'information Bayésien (*Bayes Information Criterion*)
 BRET : Transfert d'énergie de bioluminescence par résonance (*Bioluminescence Resonance Energy Transfer*)
 BSA : Albumine de sérum bovin (*Bovine Serum Albumin*)
 DAG : Diacylglycérol
 DMEM : Milieu *Eagle* modifié de Dubelcco (*Dubelcco's Modified Eagle Medium*)
 DOPC : 1,2-dioleoyl-*sn*-glycero-3-phosphocholine
 ECL : Boucle extracellulaire (*ExtraCellular Loop*)
 EDTA : Acide Éthylènediaminetétraacétique
 ELISA : Dosage immuno-enzymatique sur support solide (*Enzyme-Linked ImmunoSorbent Assay*)
 ERK : Kinase régulée de façon extracellulaire (*Extracellular Regulated Kinase*)
 FBS : Sérum bovin fœtal (*Fetal Bovine Serum*)
 FSHR : Récepteur de l'hormone stimulatrice des follicules (*Follicle-Stimulation Hormone Receptor*)
 GAP : Protéine activatrice de GTPase (*GTPase-Activation Protein*)
 GDP : Guanosine diphosphate
 GEF : Facteur d'échange de nucléotides de guanine (*Guanine nucleotide Exchange Factor*)
 GFP : Protéine fluorescente verte (*Green Fluorescent Protein*)
 GMHMM : Modèle de Markov caché avec mélange de Gaussiennes (*Gaussian Mixture Hidden Markov Model*)
 GPCR : Récepteur couplé aux protéines G (*G Protein-Coupled Receptor*)
 GRK : Kinase de récepteurs couplés aux protéines G (*G protein Receptor Kinase*)
 GTP : Guanosine triphosphate
 HA : Hémagglutinine
 HBSS : Solution de sels balancée de Hanks (*Hanks' Balanced Salt Solution*)
 HEK : Embryonnaires humaines du rein (*Human Embryonal Kidney*)
 HEPES : Acide 4-(2-hydroxyethyl)-1-piperazineéthanesulfonique
 HHM : Mécanisme d'encombrement hydrophobe (*Hydrophobic Hindering Mechanism*)
 HMM : Modèle de Markov caché (*Hidden Markov Model*)
 ICL : Boucle intracellulaire (*IntraCellular Loop*)
 mAb : Anticorps monoclonal
 MAPK : Kinase de protéines activant la mitose (*Mitosis-Activating Protein Kinase*)
 MEM : Milieu *Eagle* modifié (*Modified Eagle Medium*)
 MHBN ou MHN : Réseau majeur de ponts hydrogène (*Major H-bond Network*)
 MSM : Modèle d'états Markovien (*Markov State Model*)
 PBS : Saline tamponnée au phosphate (*Phosphate-Buffered Saline*)
 PCA : Analyse des composantes principales (*Principal Component Analysis*)
 PCCA : Analyse groupe-groupe de Perron (*Perron Cluster-Cluster Analysis*)

PCR : Réaction en chaîne avec polymérase (*Polymerase chain reaction*)
PEI : Polyéthylèneimine
PIP2 : Phosphatidylinositol 4,5-bisphosphate
PKA : Protéine kinase A
PKC : Protéine kinase C
PLC : Phospholipase C
RET : Transfert d'énergie par résonance (*Resonance Energy Transfer*)
RGS : Régulateurs de la signalisation des protéines G (*Regulators of G protein Signaling*)
SEM : Erreur standard de la moyenne (*Standard Error of the Mean*)
SPC : Charge ponctuelle simple (*Simple Point-Charge*)
TBS : Saline tamponnée au Tris (*Tris-Buffered Saline*)
tICA : Analyse des composantes indépendantes par délai temporel (*time-lagged Independent Component Analysis*)
TM : Domaine transmembranaire (*TransMembrane domain*)
TMB : 3,3',5,5'-Tétraméthylbenzidine
TP : Température pièce
WT : Type sauvage (*Wild Wtype*)

Remerciements

Je tiens à remercier Éric Marsault et Pierre Lavigne, mes co-directeurs de maîtrise, pour leur soutien et leur patience. Merci également à Martin Montagne pour ses précieux conseils sur la culture bactérienne, la PCR et la génération de mutants, à Xavier Sainsily pour m'avoir montré comment effectuer des essais de liaison par compétition, des ELISAs de surface et des essais BRET, ainsi qu'à Elie Besserer-Offroy pour m'avoir montré comment optimiser le BRET pour produire plus de résultats en moins de temps. Merci au laboratoire de Philippe Sarret pour avoir fourni l'équipement nécessaire à ces expériences. Merci finalement à Alexandre Murza et Kien Trân pour avoir fourni les ligands nécessaires aux expériences de liaison par compétition et de BRET.

Avant donc que d'écrire, apprenez à penser
Selon que notre idée est plus ou moins obscure
L'expression la suit, ou moins nette, ou plus pure
Ce que l'on conçoit bien s'énonce clairement
Et les mots pour le dire arrivent aisément
[...]
Hâtez-vous lentement et sans perdre courage
Vingt fois sur le métier remettez votre ouvrage
Polissez-le sans cesse et le repolissez
Ajoutez quelques fois et souvent effacez

- Nicolas Boileau, l'Art poétique, chant I

Introduction: biochimie des GPCRs

Les GPCRs

Fonction et intérêt

Les récepteurs couplés aux protéines G (GPCRs) forment une famille de protéines quasi-ubiquitaires chez les eucaryotes, contenant chez l'humain plus de 800 membres (Fredriksson R, Schiöth HB, 2005). Ce sont des protéines transmembranaires, situées sur la membrane cytoplasmique et responsables d'un phénomène appelé transduction du signal, soit la conversion d'un signal extracellulaire en un signal intracellulaire. Les GPCRs et leurs protéines effectrices ont une vaste gamme de fonctions associées, entre-autres aux systèmes nerveux, cardiovasculaire, osseux, immunitaire et endocrine, à l'homéostasie, à la survie cellulaire et au développement embryonnaire (Wettschureck et Offermanns, 2005; Premont et Gainetdinov, 2007).

Cette diversité de fonctions est représentative de leur nombre, de leur diversité de structure primaire et de la diversité de leurs ligands : les récepteurs humains sont regroupés en cinq familles sous la classification GRAFS, dont la famille R (Rhodopsin-like) est la plus nombreuse (Fredriksson *et al*, 2003). La diversité des structures primaires des GPCRs cache toutefois des caractéristiques communes : sept domaines transmembranaires, des interactions entre domaines transmembranaires et des résidus conservés (Venkatakrisnan *et al*, 2013) et le partage de quelques partenaires d'interaction - les protéines G, les GRKs et les arrestines - au niveau intracellulaire (Wettschureck et Offermanns, 2005; Premont et Gainetdinov, 2007; Ritter et Hall, 2009; Xie et Palmer, 2007).

Les GPCRs, de par la diversité de leurs fonctions, forment une des familles de protéines les plus importantes pour l'industrie pharmaceutique : plus de 25% des composés développés par l'industrie pharmaceutique ciblent les GPCRs (Hopkins et Groom, 2002; Fang *et al*, 2015). De plus, seuls 10% des récepteurs humains sont présentement la cible de composés développés par l'industrie pharmaceutique et près de 140 GPCRs n'ont pas encore de ligand connu (Fang *et al*, 2015). La large part de l'industrie pharmaceutique, leur vaste

gamme de fonctions ainsi que le potentiel de développement qu'ils représentent rend leur importance économique et thérapeutique indéniable.

Génétique et diversité

Les GPCRs sont regroupés en familles selon divers critères, les plus utilisés étant leur similarité de structure primaire et le type de domaine de liaison au ligand. Le premier critère permet de classifier les GPCRs en un ensemble de cinq familles, la classification GRAFS (Fredriksson *et al*, 2003), et le second critère en trois familles : A, B et C (Bridges et Lindsley, 2008). Cela dit, même au sein des familles GRAFS, les GPCRs présentent une forte diversité de structure primaire (Fredriksson *et al*, 2003).

La première famille de structures primaires de la classification GRAFS est celle des récepteurs du glutamate. Elle est caractérisée par une section N-terminale de grande taille (280-580 AAs) formant le site de liaison du ligand. Ce domaine de liaison caractéristique prend la forme de deux lobes qui se resserrent autour du ligand, d'où l'appellation de *Venus Flytrap Domain* (Fredriksson *et al*, 2003).

La seconde famille est celle des récepteurs de type rhodopsine (*Rhodopsin-like*). Il s'agit de loin de la famille de GPCRs la plus nombreuse chez l'humain. Cette famille est caractérisée par une section N-terminale plus petite et un site de liaison du ligand au moins partiellement situé entre les domaines transmembranaires. C'est de loin la famille la plus étudiée, puisque c'est elle qui présente le plus d'applications pharmaceutiques : les récepteurs adrénergiques, muscariniques, opioïdes, de chimiokines, de l'angiotensine, *etc.* font tous partie de cette famille (Fredriksson *et al*, 2003).

La troisième famille, celle des récepteurs d'adhésion, est caractérisée par une section N-terminale de longueur extrêmement variable (200-2800 AAs) et possédant des motifs d'adhésion à la matrice extracellulaire (Fredriksson *et al*, 2003).

La quatrième famille, Frizzled/Tas2, est en fait une famille double : les récepteurs de type Frizzled sont similaires aux récepteurs Tas2, mais présentent des différences majeures dans la section N-terminale. Les récepteurs Frizzled possèdent une section N-terminale de près de 200 résidus avec une série de cystéines conservées servant à la liaison avec la protéine Wnt. Les récepteurs Tas2, en revanche, ont une section N-terminale quasi-inexistante et

leur fonction est mal connue. Une de leurs fonctions supposées est la détection du goût amer (Fredriksson *et al*, 2003).

La dernière famille, celle des récepteurs de type Sécérétine, est caractérisée par une section N-terminale relativement courte (60-80 AAs) et très riche en ponts disulfure. La section N-terminale, essentielle pour la liaison du ligand de ces récepteurs, est ainsi fortement contrainte au niveau structural (Fredriksson *et al*, 2003).

Toutes ces familles de GPCRs ont des fonctions, des ligands et des façons d'interagir avec leur ligand (modes de liaison) qui leur sont propres. Il est cependant pertinent de classifier les modes de liaison en trois types principaux, ce qui mène à la classification ABC des GPCRs. Les récepteurs de classe A interagissent avec leur ligand à l'aide de leurs domaines transmembranaires, ceux de la classe B uniquement avec leur section N-terminale et ceux de la classe C avec une section N-terminale de type *Venus Flytrap Domain* (Bridges et Lindsley, 2008).

Dans le contexte de ce mémoire, tout ce qui suit sur les GPCRs traite uniquement des récepteurs de la famille Rhodopsin-like qui sont aussi dans la classe A, sauf en cas de mention contraire.

La structure des GPCRs

La structure des GPCRs est d'une importance capitale à la compréhension de leur fonction. En effet, leur rôle est de transmettre un signal de l'extérieur de la cellule à l'intérieur de la cellule sans que les deux ne communiquent de façon significative. Les GPCRs agissent ainsi comme une sorte d'antenne. C'est les détails de leur configuration structurale qui détermine comment cette antenne se comporte et quel signal extracellulaire correspond à quel signal intracellulaire.

Données actuelles

Les GPCRs sont des protéines transmembranaires situées sur la membrane cytoplasmique. Ils sont tous composés d'une région N-terminale (N-terminus ou N-term) extracellulaire, de trois boucles extracellulaires (ECL1-3), de sept domaines transmembranaires (TM1-7)

sous forme d'hélice alpha, d'une hélice alpha intracellulaire (H8), de trois boucles intracellulaires (ICL1-3) et d'une région C-terminale (C-terminus ou C-term) intracellulaire. Ceci s'applique non seulement aux récepteurs Rhodopsin-like de la classe A, mais aussi à tous les GPCRs (Venkatakrishnan *et al*, 2013).

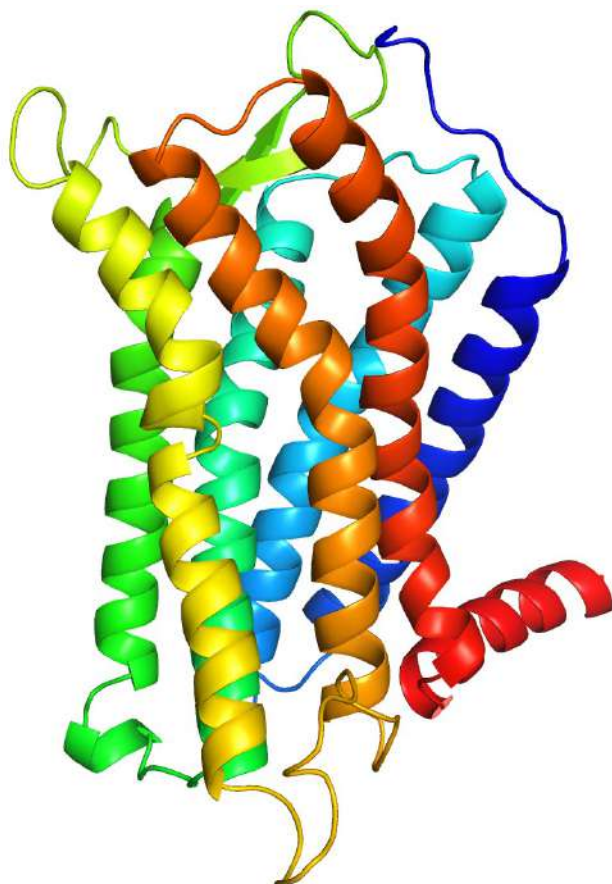


Figure 1 : Structure et régions importantes d'un GPCR.

La portion N-terminale est colorée en bleu et la portion C-terminale est colorée en rouge. Le reste est en dégradé en fonction de la position du résidu dans la structure primaire. Les domaines transmembranaires sont représentés avec les couleurs bleu foncé (TM1), bleu pâle (TM2), turquoise (TM3), vert (TM4), jaune (TM5), orange (TM6) et rouge (TM7). L'hélice 8 est située en bas à droite, en rouge rubis et la région C-terminale désordonnée n'est pas montrée ici, mais suit l'hélice 8.

Les informations existantes sur la structure des GPCRs proviennent d'un répertoire de plus en plus vaste de structures calculées à partir de résultats de cristallographie aux rayons X. Plus de 75 structures ont été publiées depuis 2000, dont plus de 70 depuis 2010 (Tableau 1), avec des ligands variés et même à quelques reprises avec des partenaires d'interaction intracellulaires (Tableau 1, en gras). Tous présentent les caractéristiques structurales décrites précédemment.

Tableau 1 : Codes PDB représentatifs de structures publiées sur chaque GPCR dont au moins une structure existe sur la PDB

* Structure calculée à partir de résultats de RMN

† : Structure calculée à partir de résultats de Cryo-EM

En gras : co-cristallisation avec une protéine effectrice (entre parenthèses)

Souligné : co-cristallisation avec un anticorps

En italique : cristallisé avec un agoniste (ou constitutivement actif), sans anticorps ni protéine effectrice

En gras contour : le récepteur n'est pas de classe A, ni rhodopsin-like

Récepteur (gène)	Codes PDB
RHO	1F88, 2Z73, 2X72 , 3PQR , 4A4M (<i>finger loop</i> d'arrestine), 4ZWJ , 5DGY , 5W0P (arrestine visuelle)
Opsine	3DQB, 4J4Q, 4PXF, 5DHG, 5WKT
ADRB1	2VT4, 2Y00-2Y04
ADRB2	2RH1, <u>3P0G</u> , <i>3PDS</i> , <u>3SN6</u> (Hétérotrimère G _s), 5D6L, <u>5JQH</u> , 5X7D
ADORA1	5UEN, 6D9H (chimère avec CHMR4) [†] (Hétérotrimère G _{i2})
ADORA2A	2YDO, 3EML, <i>3QAK</i> , 3PWH, <i>4UG2</i> , 5G53 (Mutant Gα _s), 5IU4, 5K2D, 5NLX, 5OLG, 5UIG, 5UVI, 5VRA, <i>5WF6</i> , 6AQF, <u>6GDG</u> [†] (miniG _s)
CCR2	5T1A
CCR5	4MBS
C5AR1	5O9H, 6C1Q
CXCR1	2LNL*
CXCR4	3ODU, 4RWS
CRF1R	4K5Y, 4Z9G
OPRM1	4DKL, 5C1M, <u>6DDE</u> [†] (Hétérotrimère G _{i1})
OPRD1	4EJ4, 4N6H, 4RWA
OPRK1	4DJH, <u>6B73</u>

DRD2	6CM4
DRD3	3PBL
DRD4	5WIU
AGTR1	4YAY, 4ZUD
AGTR2	5UNF, 5XJM
APLNR	<i>5VBL</i>
S1PR1	3V2Y
HRH1	3RZE
CHRM1	5CXV
CHRM2	3UON
CHRM3	4DAJ, 4U14
CHRM4	5DSG
CBR1	5U09
HTR1B	5V54
HTR2B	4NC3, 5TUD, 5TVN
HTR2C	6BQG
PAR1	3VW7
PAR2	5NDD
OPRL1	4EA3, 5DHG
SMO	4JKV, 4N4W, 4O9R, <i>4QIN</i> , 5L7D, 5V56, 6D32
GCGR	4L6R, 5EE7
GLP1R	<i>5NX2</i> , <u>5VAI</u> [†] (Hétérotrimère G _s), 5VEW, <u>6B3J</u> [†] (Hétérotrimère G _s)
GRM1	4OR2
GRM5	4O09, 5CGC, 6FFI

P2RY1	4XNV
P2RY12	4NTJ, 4PY0, 4PXZ
LPAR1	4Z34
HCRTR1	4ZJC
EDNRB	5GLH
FFAR1	4PHU
CALCR	<u>5UZ7</u> [†] (Hétérotrimère G _s)
NTSR1	4GRV, 4XEE, 5T04
CMV US28	<u>4XT1</u> , <u>5WB1</u>

Notation Ballesteros-Weinstein

Cette structure commune a mené à l'établissement d'une notation permettant le positionnement d'un résidu au sein d'un domaine transmembranaire, ce qui permet la comparaison de résidus structurellement proches dans plusieurs GPCRs en faisant fi de la similarité de structure primaire. Cette notation, appelée notation Ballesteros-Weinstein en l'honneur de ses inventeurs, assigne au résidu le plus conservé de chaque domaine transmembranaire le nombre 50 (i.e. 1.50, 2.50 ... 7.50). Les résidus situés plus près du C-term voient leur nombre augmenter et inversement pour le N-term (i.e. N-term - ... - 1.49 - 1.50 - 1.51 - ... - C-term) (Ballesteros et Weinstein, 1995). Comme ces résidus n.50 sont fortement conservés dans les différentes structures primaires disponibles des GPCRs et sont suspectés être situés au même endroit dans la structure de tous les GPCRs, les autres résidus sont ainsi associés à une région bien précise de chaque domaine transmembranaire, ce qui permet une comparaison des résidus de GPCRs différents sur une base structurale (Venkatakrishnan *et al*, 2013). Cette notation sera utilisée de façon assez fréquente plus loin dans ce mémoire, pour indiquer ou comparer le rôle dynamique de certains résidus dans la signalisation par les GPCRs.

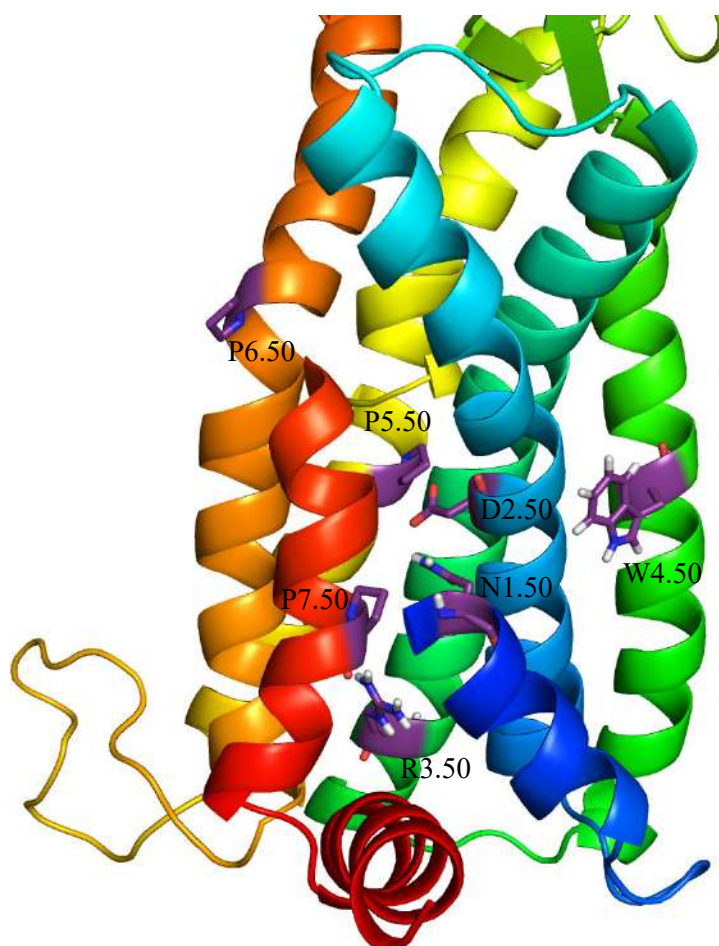


Figure 2 : Localisation des résidus n.50 sur le récepteur hAPJ.

Le même code de couleur que la figure 1 est utilisé, avec les résidus n.50 en mauve. Chaque résidu est identifié avec son type et le domaine transmembranaire auquel il appartient.

Interactions conservées entre les domaines transmembranaires

Les domaines transmembranaires des GPCRs forment un “panier” dans lequel se loge le ligand. Cette structure est conservée dans toutes les structures connues de GPCRs. Chaque domaine transmembranaire est relié aux autres par un ensemble de contacts non-covalents conservés (Tableaux 1 et 2). Ces contacts sont supposés maintenir l’intégrité structurelle des GPCRs, ainsi que de contribuer à la transduction du signal. Dans les structures connues, les contacts conservés entre le ligand et le récepteur s’effectuent avec les résidus 3.32, 3.33, 3.36, 6.48, 6.51 et 7.39. Les autres contacts entre le ligand et le récepteur sont plus variables, mais sont tout de même présents. Il advient aussi que les résidus 6.48 et 6.51 sont impliqués dans les contacts conservés entre les domaines transmembranaires, ce qui indiquerait leur rôle potentiel dans la transduction du signal (Venkatakrishnan *et al*, 2013).

Tableau 2 : Nombre de contacts conservés totaux et intracellulaires entre chaque paire de domaines transmembranaires (Venkatakrisnan <i>et al</i> , 2013).		
Interface	Contacts totaux	Contacts intracellulaires
1-2	2	2
1-7	3	2
2-3	1	1
2-7	2	2
3-4	1	0
3-5	2	1
3-6	2	1
5-6	1	1
6-7	2	0

Les sept domaines transmembranaires ont une importance particulière dans la fonction des GPCRs, mais le TM3 semble être celui le plus important. En effet, près de 50% des contacts entre domaines transmembranaires impliquent le TM3. Celui-ci est aussi impliqué dans les contacts conservés entre le ligand et le récepteur, en plus d'être lié par un pont disulfure conservé avec la boucle extracellulaire 2 et un pont-H avec la boucle intracellulaire 2 dont l'importance structurelle est avérée (Warne *et al*, 2008). Ce rôle structural et fonctionnel important est permis par l'angle important que le TM3 fait avec la membrane cytoplasmique (autour de 35°), ce qui lui permet d'interagir avec des portions variées du récepteur (Venkatakrisnan *et al*, 2013).

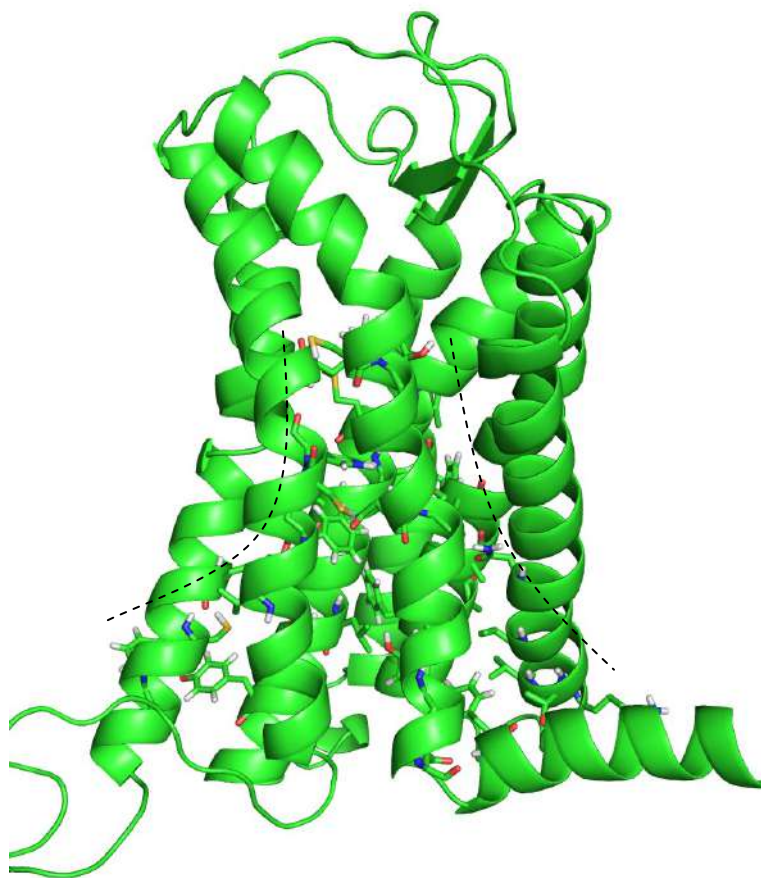


Figure 3 : Contacts non-covalents conservés chez les GPCRs

L'ensemble des contacts comprend 13 groupes distincts qui descendent du bas du site de liaison du ligand jusqu'au site de liaison des protéines G et des arrestines. Pour plus d'information sur chacun des groupes de contacts conservés, veuillez consulter Venkatakrishnan *et al*, 2013.

Domaines extracellulaires et intracellulaires

Le N-term est une région variable en forme et en structure primaire. Son rôle est de contribuer à la liaison du ligand à son récepteur (Fredriksson *et al*, 2003; Venkatakrishnan *et al*, 2013). Comme la plupart des ligands des récepteurs Rhodopsin-like sont hydrosolubles, le N-term est souvent tenu à l'écart de la pochette de liaison, qui reste ouverte. Dans certains cas, lorsque le ligand est liposoluble, le N-term contribue à la formation d'un bouchon qui empêche l'eau de circuler dans la pochette de liaison (Venkatakrishnan *et al*, 2013).

Les trois boucles extracellulaires sont les structures primaires qui relient les TM2-3 (ECL1), TM4-5 (ECL2) et TM6-7 (ECL3). Selon le récepteur, les boucles extracellulaires peuvent ou non avoir un rôle dans la liaison du ligand. L'ECL2 est la plus longue et est celle qui présente le plus de variabilité structurelle : elle peut prendre la forme d'une hélice

alpha, d'un feuillet beta à deux brins antiparallèles ou avoir une structure désordonnée. Elle couvre partiellement le site de liaison et contribue à l'affinité du récepteur pour son ligand. Il existe aussi un pont disulfure, conservé dans la plupart des structures connues de GPCRs, entre l'ECL2 et le TM3. Contrairement à l'ECL2, les boucles ECL1 et ECL3 sont petites et présentent peu de variabilité (Venkatakrishnan *et al*, 2013).

Les trois boucles intracellulaires sont les structures primaires qui relient les TM1-2 (ICL1), TM3-4 (ICL2) et TM5-6 (ICL3). L'ICL1 est en général petite, 6 AAs en moyenne, n'a pas de rôle spécifique, hormis celui d'espaceur entre les TM1-2. L'ICL2 contient une hélice alpha à un ou deux tours, ou encore une région désordonnée. Cette boucle intracellulaire est, dans certaines structures de récepteurs opioïdes ou aminergiques, reliée au TM3 par un pont-H (parfois aussi d'un pont salin) entre D3.49 et une arginine ou une tyrosine de l'ICL2 (Thompson *et al*, 2012). La mutation de Tyr141, responsable de ce pont-H dans le β 1-AR humain, a ainsi été démontrée déstabilisante pour le récepteur (Warne *et al*, 2008). Pour ce qui est de l'ICL3, les structures plus anciennes des GPCR la remplaçaient par un élément rigide (Venkatakrishnan *et al*, 2013), mais de plus en plus de structures sont disponibles avec l'ICL3 complète.

La région suivant le TM7 comprend l'hélice 8 ainsi qu'une longue région intrinsèquement désordonnée qui marque la fin de la structure primaire du GPCR (Venkatakrishnan *et al*, 2013). L'hélice 8 est située immédiatement après le TM7 et, contrairement aux TMs, elle est orientée parallèlement à la membrane. Elle contient un ou plusieurs sites de palmitoylation qui l'y accrochent (Huynh *et al*, 2009). La région désordonnée a pour rôle de servir de substrat aux GRKs, qui la phosphorylent à des endroits spécifiques, formant ainsi une sorte de « code barre » dont les motifs contribuent à déterminer la suite de la chronologie de la signalisation des GPCRs (Liggett, 2011; Sente *et al*, 2018). De plus, ces motifs de phosphorylation servent au recrutement des arrestines (Zhou *et al*, 2017).

La signalisation des GPCRs

Chronologie

La signalisation des GPCRs est un processus qui s'effectue en plusieurs étapes et implique de nombreux intermédiaires. Elle s'articule autour de quatre processus majeurs : la liaison du ligand, la phase d'activation des protéines G, la désensibilisation et l'internalisation, qui impliquent les RGS, les GRKs et les arrestines (Wettschureck et Offermanns, 2005; Premont et Gainetdinov, 2007; Ritter et Hall, 2009; Xie et Palmer, 2007).

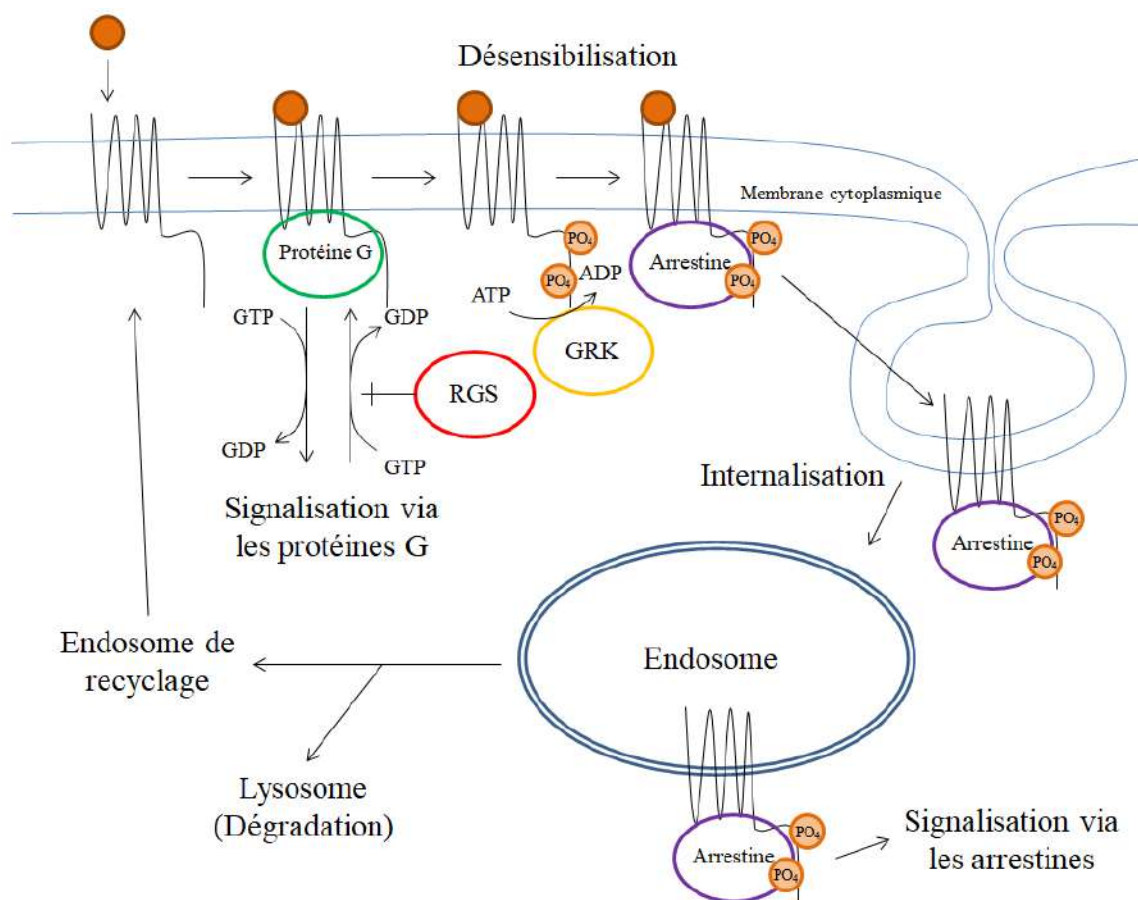


Figure 4 : Chronologie de la signalisation des GPCRs.

La liaison du ligand (en haut à gauche) initie le processus. Une fois le ligand associé au récepteur, la signalisation via les protéines G suit, puis l'action combinée des RGSs et des GRKs initie le processus de désensibilisation, qui culmine par l'attachement des arrestines au GPCR. S'ensuit l'internalisation, qui inclut la signalisation via les arrestines et, ultimement, la dégradation ou le recyclage du récepteur.

La liaison du ligand est l'étape qui déclenche la signalisation. Le processus par lequel la liaison du ligand favorise la liaison des protéines G est encore mal connu et des pistes de

solutions seront proposées plus loin. Il n'en reste pas moins que c'est suite à la liaison du ligand que l'affinité d'un GPCR pour les protéines G augmente. La phase d'activation des protéines G est caractérisée par une activité maximale des protéines G. Ce sont les premiers messagers des GPCRs et elles activent une gamme variée de cibles intracellulaires. Les principaux effecteurs en aval des protéines G (e.g. PKA, Akt, ERK, PKC) (Neves *et al*, 2002) ont des rôles biologiques très larges et l'effet global de la combinaison de chacune des voies de signalisation dépend fortement du type de récepteur, du ligand et du type cellulaire (Wettschureck et Offermanns, 2005).

La phase de désensibilisation est initiée par la phosphorylation des GPCRs par les GRKs. Cette phase est initiée en même temps que l'activation des protéines G, puisque les GRKs ont une affinité supérieure pour la forme activée des GPCRs et semblent lier les GPCRs au même endroit que les protéines G (Pitcher *et al*, 1998; Komolov *et al*, 2017). La phosphorylation des GPCRs par les GRKs contribue à de 50 à 100% de l'effet de la désensibilisation des GPCRs en augmentant l'affinité de celui-ci pour les arrestines (Pitcher *et al*, 1998). Une autre partie de la désensibilisation provient d'une rétroaction négative sur l'activité des protéines G par d'autres protéines en aval de la cascade de signalisation (Gainetdinov *et al*, 2004). La phosphorylation des GPCRs permet ensuite leur liaison avec arrestines avec une affinité de beaucoup supérieure (Gainetdinov *et al*, 2004; Moore *et al*, 2007). Ces arrestines entrent alors en compétition avec les protéines G pour la liaison aux GPCRs. En conséquence, les protéines G sont de moins en moins activées, ce qui induit une diminution de la signalisation via ces protéines (Gainetdinov *et al*, 2004; Moore *et al*, 2007).

La phase d'internalisation est initiée par la liaison du complexe arrestine-GPCR aux clathrines ou à la protéine adaptatrice AP-2, qui se lie aux clathrines. S'ensuit alors l'endocytose des GPCRs. Une fois internalisés, les arrestines continuent de stimuler certaines voies de signalisation, notamment la voie des MAPK/ERK. Un processus d'adressage s'effectue en parallèle, ce qui mène à deux destins différents : la dégradation et le recyclage. Le destin d'un GPCR est déterminé par sa portion C-terminale. Les récepteurs dits de classe A sont caractérisés par une dissociation des arrestines tôt après l'internalisation et ont tendance à être recyclés. Les récepteurs dits de classe B ont quant à

eux une affinité prolongée pour les arrestines, avec lesquelles ils colocalisent dans les endosomes, et ont tendance à être dégradés. L'échange de la région C-term entre un récepteur de classe A et un autre de classe B inverse les destins de ces GPCRs (Moore *et al*, 2007).

Le processus décrit ici est le processus dit canonique de la signalisation des GPCRs. Cela dit, il existe des preuves indiquant que le processus est plus compliqué qu'il n'y paraît. Par exemple, il a été rapporté chez certains GPCRs que leur signalisation via les protéines G était prolongée après le processus attendu de désensibilisation et d'internalisation, et que ce processus de signalisation « anormal » via les protéines G survenait chez des GPCRs internalisés. Des interactions combinées GPCR-protéine G-arrestine ont été répertoriées, formant ainsi un « super-complexe » où la protéine G continue d'exécuter sa fonction et où l'arrestine, attachée uniquement via la queue C-terminale du GPCR, permet son internalisation (Thomsen *et al*, 2016). De plus, il existe une panoplie d'autres protéines effectrices des GPCRs qui ne font pas partie du processus canonique discuté ici (Ritter et Hall, 2009).

Les protéines G

Les protéines G sont une classe de protéines hétérotrimériques qui servent comme effecteurs de la transduction du signal des GPCRs. Ce sont principalement les protéines G qui déterminent les voies de signalisation associés à l'activation des GPCRs (Neves *et al*, 2002; Moore *et al*, 2007).

La forme inactive des protéines G contient trois sous-unités : la sous-unité α , principale responsable de la régulation des effecteurs des GPCRs, et les sous-unités β et γ , qui ont une haute affinité mutuelle et peuvent être considérées comme une seule unité fonctionnelle (Hepler et Gilman, 1992) (ici appelée $G\beta\gamma$). Il existe plusieurs isoformes des sous-unités β et γ , mais la différence entre ces isoformes sort du contexte de cet ouvrage. La sous-unité α est responsable de l'activation des effecteurs des protéines G et possède une activité GTPase, c'est-à-dire la capacité d'hydrolyser la GTP en GDP. L'unité $G\beta\gamma$ est associée à l'activation de certains effecteurs, notamment l'activation de canaux potassiques (Whorton

et MacKinnon, 2013), et le relâchement de calcium intracellulaire dans le cytoplasme (Neves *et al*, 2002). L'activation du complexe $G\alpha\beta\gamma$ est provoquée directement par les GPCRs, qui agissent comme GEFs au niveau de la sous-unité $G\alpha$ en favorisant l'échange d'une GDP pour une GTP. Cet échange réduit l'affinité de $G\alpha$ pour $G\beta\gamma$, ce qui induit la dissociation des deux unités et les rend capables d'exercer leurs effets dans la cellule (Hepler et Gilman, 1992).

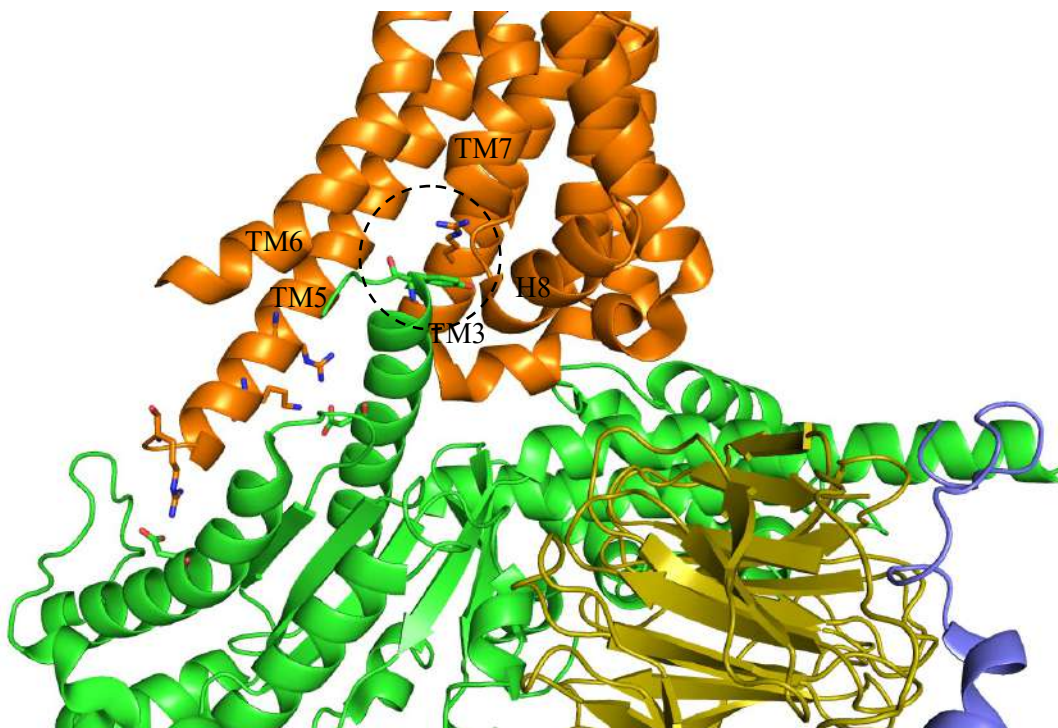


Figure 5 : Interaction entre une protéine G et un GPCR.

Cette interaction implique principalement l'ICL2 et l'ICL3 et principalement des interactions hydrophobes, à l'exception de deux résidus de l'ICL1, de quelques interactions ioniques, ainsi qu'une interaction cation-aromatique entre le résidu R3.50 et une tyrosine de la protéine G (encerclé).

La sous-unité $G\alpha$ est la sous-unité qui présente le plus de variabilité en termes d'activité intracellulaire. Il existe plusieurs sous-unités $G\alpha$ différentes, regroupées sous quatre ou cinq familles : $G\alpha_s$, $G\alpha_i$, $G\alpha_q$ et $G\alpha_{12/13}$ (Wettschureck et Offermanns, 2005; Neves *et al*, 2002; Hepler et Gilman, 1992). Chaque famille de protéines G est associée à des effecteurs différents ou des effets différents sur leurs effecteurs communs, malgré quelques caractéristiques communes (Neves *et al*, 2002). L'activité de la sous-unité $G\alpha$, de par son

activité GTPase, n'est que transitoire : éventuellement, la GTP est hydrolysée en GDP et $G\alpha$ regagne son affinité pour $G\beta\gamma$.

La famille des sous-unités $G\alpha_s$ ($G\alpha$ “stimulantes”) comprend seulement trois membres, dont le principal est $G\alpha_s$, présent de façon ubiquitaire dans l'organisme. Les sous-unités $G\alpha_s$ ont pour caractéristique principale d'activer directement l'adénylate cyclase, ce qui provoque l'accumulation d'AMPc dans la cellule, et donc indirectement l'activation de la PKA. Les effets de la PKA sont très variés et sont associés à de nombreux effets physiologiques de l'activation des $G\alpha_s$ (Wettschureck et Offermanns, 2005; Thompson *et al*, 2012; Neves *et al*, 2002).

La famille des sous-unités $G\alpha_i$ ($G\alpha$ “inhibitrices”) comprend huit membres, dont les plus importants sont les $G\alpha_{i1}$ à $G\alpha_{i3}$ et $G\alpha_o$ (Wettschureck et Offermanns, 2005). L'activité des sous-unités $G\alpha_i$ est davantage associée à ses partenaires $G\beta\gamma$ par rapport aux autres sous-unités $G\alpha$ (Wettschureck et Offermanns, 2005; Neves SR *et al*, 2002). En effet, l'activité de $G\alpha_o$ est suspectée de découler uniquement de la libération de l'unité $G\beta\gamma$, contrairement aux autres sous-unités $G\alpha$. L'activité des $G\alpha_i$ a pour caractéristique principale l'inhibition de l'adénylate cyclase, soit l'inverse des $G\alpha_s$ (Wettschureck et Offermanns, 2005).

La famille des sous-unités $G\alpha_q$ comprend quatre membres, dont les plus importants sont $G\alpha_q$ et $G\alpha_{11}$. L'activité des sous-unités $G\alpha_q$ est principalement associée au relâchement de calcium intracellulaire, par le biais de l'activation des PLC- β . L'activité des PLC- β provoque le relâchement d'IP3, ce qui active des canaux calciques sensibles à l'IP3 et permet le relâchement du calcium du réticulum endoplasmique vers le cytosol (Ferris et Snyder, 1992). Le calcium a entre autres pour partenaires la calmoduline, dont les fonctions intracellulaires sont très vastes (Cheung, 1980). Un autre effet de l'activation de la PLC- β est le relâchement de DAG, ce qui active la PKC (Wettschureck et Offermanns, 2005). Comme la calmoduline, la PKC est associée à une vaste gamme d'effets cellulaires (Berman et Gilman, 1998).

La dernière famille, celles des $G\alpha_{12/13}$, comprend seulement les membres $G\alpha_{12}$ et $G\alpha_{13}$. L'effecteur principal des $G\alpha_{12/13}$ est la protéine RhoA, impliquée dans la régulation des structures cellulaires constituées d'actine et est suspectée de contribuer à la signalisation

des cadhérines en favorisant le relâchement de la β -caténine (Wettschureck et Offermanns, 2005).

Les RGSs

Les RGSs forment une famille de GAPs responsables de l'amplification de l'activité GTPase des sous-unités $G\alpha$. En favorisant la transformation de la GTP en GDP, les RGS contribuent au processus de désensibilisation des GPCRs (Berman et Gilman, 1998; Xie et Palmer, 2007) en augmentant l'activité hydrolytique de la sous-unité $G\alpha$.

Les RGSs ne possèdent qu'une seule sous-unité et partagent tous un domaine commun : le domaine RGS, principal responsable de l'activité des RGSs (Xie et Palmer, 2007). La nature structurelle de leur fonction réside dans l'affinité du domaine RGS pour la région catalytique des $G\alpha$. En se liant à proximité du site actif, elles sont en mesure d'interagir avec les résidus et l'eau responsables de l'hydrolyse de la GTP et sont suspectées de stabiliser l'état de transition de la réaction d'hydrolyse. De ce fait, elles peuvent accélérer la vitesse de catalyse des $G\alpha$ de jusqu'à 100X.

Les RGSs ont de particulier que leur activité GAP ne semble affecter que les sous-unités $G\alpha$ des familles $G\alpha_i$ et $G\alpha_q$, en plus d'avoir une sélectivité prononcée pour certains types de GPCRs. De plus, leur expression est régulée dans le temps et l'espace, ce qui amène une grande variabilité dans leur activité inhibitrice (Xie et Palmer, 2007).

Les GRKs

Les GRKs forment une famille de kinases dont la fonction est double : elles contiennent un domaine d'homologie RGS, qui possède des fonctions similaires aux RGS, ainsi qu'un domaine kinase, qui phosphoryle les GPCRs (Xie et Palmer, 2007) et augmente ainsi leur affinité pour les arrestines (Moore *et al*, 2007). De par leurs fonctions, les GRKs contribuent à de 40 à 60% de l'effet de phosphorylation des GPCRs, ainsi qu'à de 50 à 100% de l'effet de désensibilisation, selon le type cellulaire utilisé pour l'évaluer (Pitcher *et al*, 1998).

Les GRKs ne possèdent qu'une seule sous-unité, dont la portion N-terminale contient le domaine d'homologie RGS et la portion C-terminale contient des sites d'interaction variés. Le domaine d'homologie RGS n'est pas un RGS à proprement dit, puisqu'il est moins efficace qu'un RGS classique et son mode de fonctionnement semble aussi impliquer directement les GPCRs comme activateurs, ainsi que l'inhibition des effecteurs des protéines G (Ribas *et al*, 2007).

Les GRKs sont divisées en trois sous-familles : les GRKs visuelles (GRK1,7), les GRKs 2 et 3 et la famille "GRK4" (GRK4-6). La subdivision entre les deux sous-familles non-visuelles de GRKs est associée principalement à des différences dans la région C-terminale. Les GRKs 2 et 3 possèdent une longue région C-terminale qui comprend entre-autres un domaine d'homologie de la Pleckstrine qui leur permet d'interagir avec les membranes (par le PIP2) et une grande variété de protéines, comme la cavéoline, la calmoduline, Akt, G $\beta\gamma$ et PI3K. De plus, la région C-terminale de GRK2 (la plus étudiée entre GRK2 et GRK3) est capable d'interagir avec les clathrines et jouerait donc possiblement un rôle dans l'internalisation des GPCRs. La GRK2 est aussi capable d'interagir avec G $\alpha_q/11$, G $\beta\gamma$ et l' α -actinine avec sa région N-terminale. La famille GRK4, elle, possède moins de partenaires d'interaction. GRK5, la plus étudiée de la famille, possède deux sites d'interaction avec le PIP2 et peut interagir avec la cavéoline et la calmoduline, ainsi qu'avec l' α -actinine. Elle a aussi deux sites d'autophosphorylation non-répertoriés chez GRK2 (Ribas *et al*, 2007).

Les GRKs ont pour substrat la queue C-terminale des GPCRs. En effet, l'ajout d'un peptide dont la structure primaire correspond à la queue C-terminale d'un GPCR réduit la constante d'affinité des GRKs pour les GPCRs à 1/100 à 1/100 000 de la normale sans changer la vitesse maximale de réaction, ce qui indique qu'il s'agit d'une inhibition compétitive. Les GRKs sont aussi activés de façon allostérique par les GPCRs : les GRKs sont sensibles au mastoparan, un peptide qui active aussi les protéines G, et le clivage enzymatique de l'ICL3 des GPCRs supprime l'activité des GRKs. Le recrutement des GRKs à la membrane, par l'intermédiaire de G $\beta\gamma$ et PIP2 pour GRK2 et PIP2 seulement

pour GRK5, est aussi essentiel à leur bon fonctionnement (Pitcher *et al*, 1998). Plus récemment, il a été démontré lors d'une étude de l'interaction entre le récepteur β_2 -adrénergique (β_2 -AR) et GRK5 que la liaison fonctionnelle de GRK5 avec le récepteur dépend bien de la présence de PIP2, mais aussi de la présence d'un agoniste (Komolov *et al*, 2017).

Les arrestines

Les arrestines sont une famille de protéines dont les fonctions sont multiples : elles servent d'agents de désensibilisation et d'internalisation des GPCRs, contribuent à la détermination du destin (recyclage ou dégradation) des GPCRs et sont impliquées dans la signalisation de certaines voies en aval des GPCRs (Moore *et al*, 2007).

Les arrestines sont constituées d'une seule sous-unité et leur structure consiste en un amalgame de deux domaines en feuillets β joints par une interface à moitié polaire chargée et à moitié hydrophobe (Han *et al*, 2001). Cette interface permet aux deux domaines de glisser l'un sur l'autre pour passer de la forme inactive à la forme active (Shukla *et al*, 2013). Les arrestines ont une préférence marquée pour les GPCRs phosphorylés par la GRK. Cette préférence est expliquée par la présence de résidus basiques dans le domaine N-terminal, là où se lie la queue C-terminale phosphorylée des GPCRs (Shukla *et al*, 2013), et cette liaison s'effectue par le déplacement de la propre queue C-terminale des arrestines, qui agit comme inhibiteur de sa fonction (Palcewski *et al*, 1994; Latorraca *et al*, 2018).

Une autre interaction importante est l'interaction, dans l'interface inter-domaine des arrestines, entre une arginine (R169 pour la β -arrestine 1 et R170 pour la β -arrestine 2) et un aspartate (D290 pour la β -arrestine 1 et la β -arrestine 2), qui se brise lors de l'activation des arrestines (Han *et al*, 2001; Shukla *et al*, 2013). La mutation R169D (ou R170D) et la mutation D290R facilitent toutes deux la liaison des arrestines aux GPCRs non-phosphorylés (Moore *et al*, 2007; Han *et al*, 2001).

Les arrestines sont impliquées dans la désensibilisation des GPCRs et dans toutes les étapes de transit des GPCRs qui suivent. Lors de l'activation et, plus tard, de la phosphorylation

des GPCRs, les arrestines, normalement situées partout dans le cytoplasme (Scott *et al*, 2002), se concentrent autour de la membrane cytoplasmique en raison de leur affinité accrue pour les GPCRs actifs et phosphorylés. Le rôle des arrestines dans la désensibilisation des GPCRs est expliqué par le fait qu'elles lient les récepteurs dans des régions semblables à celles des protéines G, via une région appelée *finger loop*. Elles préviennent donc la liaison des protéines G aux GPCRs et empêche ainsi leur activation (Luttrell *et al*, 2002; Kang *et al*, 2015). Ceci dit, même si la *finger loop* contribue à l'interaction arrestine-GPCR, l'adoption de sa conformation liée au récepteur ne semble pas essentielle pour l'activation de l'arrestine. Ce serait plutôt l'interaction entre l'arrestine et l'ICL2 et l'ICL3 qui en seraient responsables, en plus du C-terminal phosphorylé du GPCR (Latorraca *et al*, 2018).

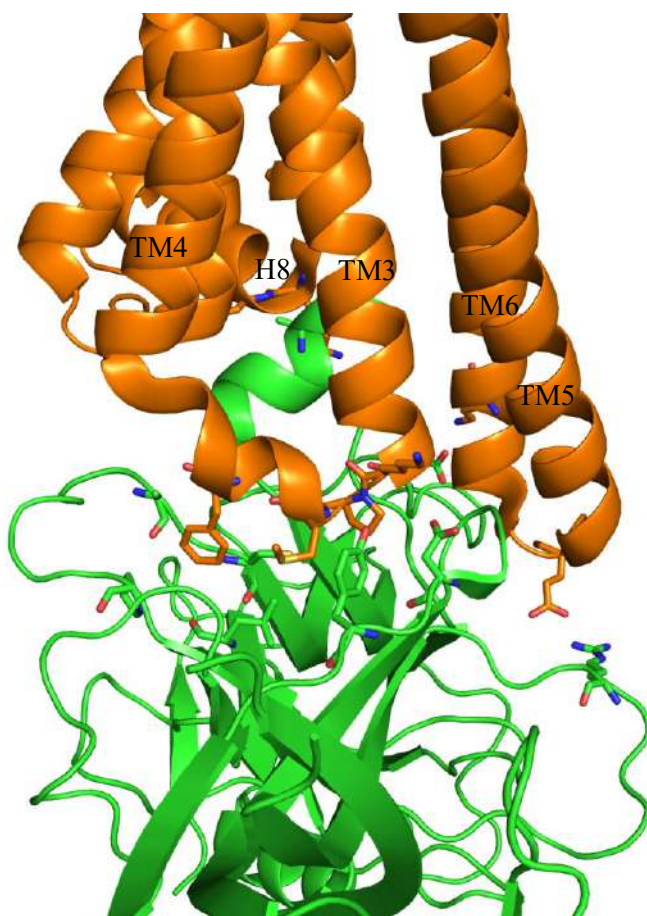


Figure 6 : Interaction entre une arrestine et un GPCR.

L'interaction implique principalement des résidus de l'ICL2 et de l'ICL3, ainsi que quelques résidus du bas du TM2 et du TM7 et du début de l'hélice 8. L'interaction avec la portion C-terminale n'est pas montrée dans cette structure, mais néanmoins présente.

Les arrestines ont des sites de liaison aux clathrines et à la protéine AP2, toutes deux localisées dans les *clathrin-coated pits*, ce qui favorise l'internalisation des GPCRs (Moore *et al*, 2007). Le rôle des arrestines en signalisation est plus indirect que celui des protéines G, mais elles servent au recrutement des protéines de la famille Src et d'une multitude d'autres protéines aux GPCRs, ainsi que d'échafaudage pour les protéines de la voie MAPK/ERK. La signalisation dépendante des arrestines a aussi de particulier qu'elle continue après l'internalisation des GPCRs (Luttrell et Gesty-Palmer, 2010). Après l'internalisation des GPCRs, les arrestines sont aussi associées au destin de chaque GPCR. La différence du comportement des GPCRs vis-à-vis les arrestines est assez importante pour les séparer en deux classes, correspondant aux deux classes de recyclage des GPCRs. Il existe ainsi des GPCRs dits de recyclage classe A, qui lient préférentiellement la β -arrestine 2, se dissocient des arrestines tôt après l'endocytose et ont tendance à être recyclés, et des GPCRs de recyclage classe B, qui lient préférentiellement la β -arrestine 1, restent liés aux arrestines longtemps après l'endocytose et ont tendance à être dégradés (Moore *et al*, 2007).

Étude de la fonction des GPCRs

Descripteurs recherchés du comportement des GPCRs et de leurs ligands

Les essais fonctionnels sur les GPCRs ont pour objectif de caractériser leur comportement dans le plus grand détail avec le moins d'essais possible. Parmi les principaux éléments recherchés dans les essais sur GPCRs, il y a :

- Le taux d'expression du récepteur à la surface de la cellule
- La capacité du récepteur à interagir avec un ligand
- La capacité du récepteur à recruter ses protéines effectrices avec ou sans le ligand
- La capacité du récepteur à activer ses protéines effectrices et provoquer une réponse cellulaire
- L'évolution de ces différentes mesures dans le temps, sous différentes conditions.

Il existe une grande variété d'autres essais fonctionnels dont il ne sera pas question ici, puisqu'ils excèdent la portée de cet ouvrage. Les essais fonctionnels d'intérêt pour ce travail seront cependant abordés plus en détail.

Liaison du ligand au récepteur et essais de liaison par saturation

La détermination de la liaison spécifique maximale et de la constante de dissociation d'un ligand à un récepteur s'effectue principalement par essai de saturation du récepteur avec un ligand radioactif (dit ligand chaud).

Cet essai se base sur le fait que la liaison entre le ligand et le récepteur est réversible, ce qui permet d'atteindre un équilibre entre les formes liées (L^*R) et non-liées (R) du récepteur. La concentration des complexes ligand-récepteur est donc évaluée en exposant une quantité constante de récepteurs à une concentration croissante de ligand chaud (L^*) :

$$L^*R \rightleftharpoons L^* + R$$

$$K_D = \frac{[L^*][R]}{[L^*R]}$$

Considérant une concentration de ligand suffisamment grande pour que la concentration en récepteurs soit négligeable (ligand en excès) et une concentration à peu près constante de récepteurs fonctionnels (i.e. la dénaturation du récepteur est ignorée) dans la préparation, on obtient :

$$[L^*]_{tot} = [L^*] + [L^*R] \approx [L^*]$$

$$[R]_{tot} = [R] + [L^*R] \approx cste$$

Considérant que l'expérience est conçue pour isoler les complexes ligand-récepteur, que la radioactivité non-spécifique est connue et retirée du compte et que la désintégration du radioisotope s'effectue de façon à peu près constante tout le long de l'expérience, il est possible d'affirmer que le compte des désintégrations N est proportionnel au nombre de complexes ligand-récepteur via la constante cinétique k_{rad} appropriée :

$$N(t, [L^*R]) \approx k_{rad}[L^*R]t$$

$$\frac{\partial}{\partial t} N(t, [L^*R]) \approx k_{rad}[L^*R]$$

En regroupant toutes ces informations, la courbe des comptes de désintégration suit une équation similaire à l'équation de Michaelis-Menten de la cinétique enzymatique, avec K_D la constante de dissociation du ligand :

$$(1) \quad \frac{\partial}{\partial t} N(t, [L^*]) \approx \frac{k_{rad}[R]_{tot}[L^*]}{K_D + [L^*]}$$

$$K_D = \frac{[L^*][R]}{[L^*R]}$$

La quantité de récepteur est normalisée par rapport à la concentration totale de protéines ($[P]_{tot}$) de l'échantillon pour donner le B_{max} , donné en moles de récepteur par gramme de protéines ou l'équivalent :

$$B_{max} = \frac{[R]_{tot}}{[P]_{tot}}$$

Pour mieux discerner les différentes variables cette équation, la régression est effectuée en fonction de la concentration du ligand radioactif sous une forme semi-logarithmique :

$$\frac{\partial}{\partial t} N(t, x) = k_{rad}[R]_{tot} \frac{1}{1 + 10^{\log K_D - x}}, x = \log[L^*]$$

Ce type de courbe admet un point d'inflexion en $x = \log K_D$:

$$\frac{\partial}{\partial x} \frac{\partial}{\partial t} N(t, x) \propto \frac{10^{x - \log K_D}}{(1 + 10^{x - \log K_D})^2}$$

$$\frac{\partial^2}{\partial x^2} \frac{\partial}{\partial t} N(t, x) \propto \frac{10^{x - \log K_D} (1 - 10^{2(x - \log K_D)})}{(1 + 10^{x - \log K_D})^4}$$

$$\frac{\partial^2}{\partial x^2} \frac{\partial}{\partial t} N(t, \log K_D) \propto \frac{1(1-1)}{(1+1)^4} = 0$$

Essais fonctionnels de la signalisation des GPCRs : le modèle opérationnel

Les essais fonctionnels sur les GPCRs présentent des défis qui leurs sont propres si on les compare aux essais de liaison. Bien qu'ils ne demandent aucun ligand radioactif, le défi associé aux essais fonctionnels réside plutôt dans la signification biologique de leurs résultats. En règle générale, plus un essai fonctionnel mesure directement l'activation d'une protéine effectrice par un GPCR, moins son résultat peut être associé avec certitude à une réponse cellulaire, et plus l'essai mesure l'effet de l'activation d'un GPCR au niveau des effecteurs en aval des cascades de signalisation, moins il peut être associé directement et avec certitude à l'activation d'un GPCR ou à une de ses protéines effectrices. Ainsi, la première approche est réductionniste, mais directe et la deuxième est complète, mais confuse.

Dans les deux cas, l'approche la plus utilisée pour aborder la relation entre le récepteur et sa fonction est le modèle opérationnel (Black et Leff, 1983). Ce modèle considère que la réponse E (pour effet) d'un effecteur (E^0 inactif, E^* activé) par rapport à la forme du récepteur liée au ligand est rectangulaire hyperbolique (i.e. similaire à un processus de Michaelis-Menten) :



$$E = \frac{E_{max}[LR]}{K_e + [LR]}$$

$$K_e = \frac{[LR][E^0]}{[LRE^0]}$$

Alors comme la concentration du complexe ligand-récepteur suit elle-même une relation rectangulaire hyperbolique avec la concentration du ligand :

$$[LR] = \frac{[R]_{tot}[L]}{K_D + [L]}$$

$$E = \frac{E_{max} \frac{[R]_{tot}[L]}{K_D + [L]}}{K_e + \frac{[R]_{tot}[L]}{K_D + [L]}} = \frac{E_{max}[R]_{tot}[L]}{K_e K_D + (K_e + [R]_{tot})[L]}$$

La grandeur appelée « coefficient de transduction », puisqu'elle témoigne de l'efficacité du récepteur à provoquer l'effet étudié, est alors définie :

$$\tau = \frac{[R]_{tot}}{K_e}$$

$$(2) \quad E = \frac{E_{max}\tau[L]}{K_D + (1 + \tau)[L]}$$

Ainsi :

$$(2) \quad E_{\infty} = \lim_{[L] \rightarrow \infty} \frac{E_{max}\tau[L]}{K_D + (1 + \tau)[L]} = E_{max} \frac{\tau}{(1 + \tau)}$$

$$\frac{E_{max}\tau EC_{50}}{K_D + (1 + \tau)EC_{50}} = \frac{E_{max}}{2} \frac{\tau}{(1 + \tau)}$$

$$\Rightarrow E_{max}\tau EC_{50} = \frac{E_{max}}{2} \frac{\tau(K_D + (1 + \tau)EC_{50})}{(1 + \tau)}$$

$$\Rightarrow E_{max}\tau EC_{50} = \frac{E_{max}}{2} \frac{\tau K_D}{1 + \tau} + \frac{E_{max}}{2} \tau EC_{50}$$

$$\Rightarrow EC_{50} = \frac{K_D}{1 + \tau}$$

Cette représentation de la relation entre la liaison du ligand au récepteur et de la fonction de ce récepteur est fréquemment utilisée dans un contexte de comparaison de plusieurs effets

du récepteur. La comparaison des τ respectifs à chaque effet étudié permet de mesurer le biais du récepteur pour un ou l'autre des effets. Il s'agit d'une façon de mesurer le biais de signalisation, dont il sera discuté plus tard.

Cela dit, le modèle opérationnel tel que décrit ci-dessus ne considère pas le fait que l'activité du récepteur peut aussi dépendre de ses formes non-liées au ligand. Autrement dit, l'activité basale du récepteur est ignorée. Dans un contexte où plusieurs ligands différents sont exposés au même récepteur, il est raisonnable d'exclure cette activité basale du calcul, mais dans un contexte où différents récepteurs (e.g. mutants d'un même récepteur) sont exposés à différents ligands, il est nécessaire d'inclure l'activité basale dans le calcul. Dans ce cas, le modèle opérationnel prend plutôt la forme suivante (Slack et Hall, 2012; Zhou *et al*, 2019) :

$$E = \frac{E_{max}(\alpha[R] + [LR])}{K_e + \alpha[R] + [LR]}$$

Le paramètre α représente ici la facilité qu'a le récepteur à provoquer l'effet sans contribution du ligand par rapport à avec contribution du ligand. Un essai effectué sans le récepteur donne le seuil minimal donne le niveau minimal de l'effet observé (E_{min}).

En considérant que $K_D = \frac{[L][R]}{[LR]}$ et que $[R]_{tot} = [R] + [LR]$, on a :

$$E = \frac{E_{max}[R]_{tot}(\alpha K_D + [L])}{K_D(K_e + \alpha[R]_{tot}) + (K_e + [R]_{tot})[L]}$$

En utilisant la même définition du τ que dans le modèle opérationnel de base, on a :

$$(3.1) \quad E = \frac{E_{max}\tau(\alpha K_D + [L])}{K_D(1 + \alpha\tau) + (1 + \tau)[L]}$$

Le système admet maintenant une activité minimale en absence de ligand et les autres paramètres sont modifiés en conséquence. La résultante est un système de trois équations et trois inconnues (α , τ et E_{max}), avec E_0 , E_∞ , K_D et EC_{50} connus :

$$(3.2) \quad E_0 = E_{rec} = E_{max} \frac{\alpha\tau}{(1 + \alpha\tau)}$$

$$E_\infty = \lim_{[L] \rightarrow \infty} \frac{E_{max}\tau(\alpha K_D + [L])}{K_D(\tau + 1) + (\tau + \alpha)[L]} = E_{max} \frac{\tau}{(1 + \tau)}$$

$$\begin{aligned} \frac{\alpha - 1}{(1 + \tau)} K_D &= \frac{\alpha - 1}{(1 + \alpha\tau)} EC_{50} \\ \Rightarrow \frac{K_D}{EC_{50}} &= \frac{(1 + \tau)}{(1 + \alpha\tau)}, \alpha \neq 1 \end{aligned}$$

Le problème est maintenant que τ n'est plus une mesure suffisante pour témoigner de l'efficacité d'un récepteur à provoquer l'effet étudié puisque α doit aussi être considéré. De plus, la présence d'autant de paramètres inconnus dans le modèle rend plus difficile et moins exacte son estimation. Ceci dit, le modèle inclut l'existence d'une activité constitutive ($0 < \alpha \leq 1$) et admet ainsi aussi l'existence d'agonistes inverses ($\alpha > 1$), ce qui le rend plus réaliste.

La conclusion est donc que la plupart des essais fonctionnels ou de liaison sur GPCRs s'analysent exactement de la même façon tant que les hypothèses de départ sont respectées : les formes actives de la plupart des effecteurs étudiés du récepteur suivent une relation hyperbolique par rapport à la concentration du ligand, et par conséquent, ces essais peuvent être interprétés de la même façon, le plus souvent sous leur forme semi-logarithmique:

$$E = E_0 + \frac{(E_\infty - E_0)10^{\log[L]}}{10^{\log(EC_{50})} + 10^{\log[L]}} = E_0 + (E_\infty - E_0) \frac{1}{1 + 10^{\log EC_{50} - x}}, x = \log[L]$$

En assumant que le bruit de fond récepteur-indépendant a été retiré, les trois mesures d'intérêt sont ici l'effet maximal observé (E_{∞}), l'effet minimal observé (E_0) et la concentration effectrice à 50% (EC_{50}) : l' E_{∞} donne l'efficacité du système ligand-récepteur-effecteur à provoquer l'effet biologique observé, l' E_0 donne l'activité basale du système et l' EC_{50} donne la concentration requise pour atteindre la moitié de l'effet ligand-dépendant. L' EC_{50} représente aussi la puissance du ligand, qui indique la capacité d'une petite concentration de ligand à provoquer un effet significatif. Plus un ligand est puissant, moins il en faut pour provoquer un effet et plus il est efficace, plus la réponse qu'il provoque est importante (E_{∞} plus grand). Dans le graphique semi-logarithmique, l' EC_{50} correspond au point d'inflexion de la courbe, tandis que l' E_{∞} est le maximum de la courbe.

Il peut aussi arriver que le processus de réponse suive plutôt un processus décrit par l'équation de Hill :

$$E = \frac{E_{max}(\alpha[R] + [LR])^n}{K_e + (\alpha[R] + [LR])^n}$$

Dans ce cas, le modèle est légèrement différent, mais le traitement de la signalisation biaisée est similaire. Pour obtenir un traitement plus détaillé de ce type de processus, veuillez vous référer à Black et Leff, 1983, ainsi que Slack et Hall, 2012).

Le modèle opérationnel a aussi comme désavantage qu'il masque un aspect important de la signalisation des GPCR. En effet, simplement en considérant qu'une succession de deux processus hyperboliques reste hyperbolique (Black et Leff, 1983), il est possible de constater que les coefficients de la fonction résultante dépendent des coefficients des deux équations initiales. Cela signifie que, même si on simplifie à l'extrême le processus de signalisation et qu'on assume une série de processus hyperboliques, l' E_{∞} , l' E_0 et l' EC_{50} de l'activation d'un effecteur dans une cascade de signalisation dépendent des constantes catalytiques, d'équilibre, d'inactivation, des constantes d'affinité et des concentrations de tous les effecteurs précédents de la cascade, jusqu'au récepteur. Ainsi, il est fort probable que l'effet d'un GPCR varie de façon extrême d'une cellule à l'autre, d'un type de cellule à

l'autre, d'un animal à un autre et même d'un environnement à un autre (e.g. composés volatils (parfums, odeurs, phéromones), température, courants d'air, alimentation, hydratation ...), selon le niveau d'expression de toutes les protéines de la cascade de signalisation. Il est d'ailleurs certain que plus le phénomène observé se trouve en aval d'une cascade de signalisation, plus l'observation est contaminée par l'effet d'autres protéines effectrices que celles directement activées par le GPCR étudié. Ce phénomène sera discuté un peu plus en détail dans la prochaine section.

La signalisation biaisée

Signalisation et effets physiologiques : efficacité pluridimensionnelle

Comme les GPCRs sont capables d'interagir avec une multitude de protéines effectrices différentes, il est raisonnable de supposer que certaines de ces protéines, ou combinaisons de ces protéines, sont en mesure de causer spécifiquement les effets physiologiques observés suite à l'exposition d'un GPCR à un de ses ligands. En effet, en particulier pour les récepteurs opioïdes (Bohn *et al*, 2000; Raehal *et al*, 2005; Manglik *et al*, 2016), cette hypothèse est appuyée par des preuves de plus en plus substantielles. L'idée de développer des ligands d'un récepteur qui serait capable de provoquer les effets thérapeutiques désirés (e.g. analgésie) sans provoquer les effets indésirables (e.g. tolérance et addiction) présente un potentiel thérapeutique (et commercial) extrêmement important. Considérant la crise actuelle en lien avec l'addiction aux opioïdes (Kolodny *et al*, 2015), une réalisation de ce concept est aussi d'intérêt public.

Ce raisonnement sur la relation entre l'interaction d'un GPCR avec une protéine effectrice spécifique et un effet physiologique spécifique a conduit à l'introduction d'un concept d'efficacité pluridimensionnelle (Kenakin, 2011), où la capacité d'un GPCR à provoquer les différents effets observés sont traités de façon indépendante. Dans le contexte où, pour deux ligands différents, les activités ligand-dépendantes d'un récepteur sont comparées, cela rend possible la mise en relation des différents τ associés aux différents effets observés pour les différents ligands d'un récepteur, qui mène à une expression quantitative du phénomène d'efficacité pluridimensionnelle, sous la forme d'une mesure du biais de

signalisation associé à un ligand par rapport à un ligand de référence (Rajagopal *et al*, 2011; Slack et Hall, 2012; Kenakin et Christopoulos, 2013). Ceci dit, deux limitations importantes au concept de biais de signalisation sont à considérer lorsqu'on observe le biais ligand-dépendant : le biais intrinsèque au système dans lequel le biais est observé (biais systémique) et le biais associé aux conditions expérimentales de l'observation (biais observationnel).

Le biais systémique

La plupart des méthodes associées à l'observation *in vitro* de l'efficacité pluridimensionnelle sont réalisées sur des systèmes de cellules immortalisées (e.g. HEK293) contraintes à la surexpression d'ADN recombinant, par exemple du récepteur et une de ses protéines effectrices, ou d'une protéine rapporteuse d'une réponse cellulaire. Ainsi, le niveau d'expression des différents éléments d'une cascade de signalisation ne ressemble souvent en rien à ce qui est observé *in vivo*. Ce type de système exclut aussi la présence de tissus biologiques, d'organe, et tout ce qui a trait à leur organisation spatiale, leurs interactions, leur physiologie, etc. (Gundry *et al*, 2017). Ce type de biais est inévitable et une vérification des résultats obtenus *in vitro* en conditions *in vivo* dans un modèle animal est donc de mise pour confirmer l'effet observé. De plus, même *in vivo* dans un modèle animal, ce type de biais existe encore par rapport à des essais chez l'humain (van der Worp *et al*, 2010).

Le biais observationnel

Au sein d'un même système, différentes conditions physico-chimiques peuvent avoir un impact sur l'effet observé. Les concentrations de différents ions métalliques (e.g. Na^+ , K^+ , Ca^{2+}), par exemple, peuvent moduler le K_D de certains ligands, ainsi qu'avoir un impact sur l'équilibre des populations d'un GPCR, donc nécessairement sur les E_0 , E_∞ et EC_{50} observés (Gundry *et al*, 2017; Ye *et al*, 2018; Vsevolod *et al*, 2014). De même, une température ou un temps d'incubation différents peuvent même aller jusqu'à inverser le biais observé (Herenbrink *et al*, 2016). Ce type de biais est plus facilement contrôlé, mais

doit tout de même être pris en considération lors de la conception d'une expérience et la comparaison de différents résultats disponibles dans la littérature.

Le biais ligand-dépendant

Une fois que les biais systémique et observationnel sont pris en compte, il ne reste que le biais de signalisation ligand-dépendant. Dans un même système et dans les mêmes conditions expérimentales, ce biais survient lorsqu'un récepteur est exposé à un ligand test et que la réponse observée diffère substantiellement de la réponse d'un ligand de référence. Dans ce cas, le biais de signalisation ligand-dépendant du ligand test par rapport au ligand de référence peut être calculé de deux façons. La première utilise le modèle équiactif et la seconde utilise le modèle opérationnel (Rajagopal *et al*, 2011; Kenakin et Christopoulos, 2013; Kenakin, 2011).

Le modèle équiactif pour la quantification du biais ligand-dépendant utilise la relation linéaire entre l'inverse des concentrations du ligand test et du ligand de référence pour obtenir le même effet observé (Rajagopal *et al*, 2011) :

$$E_{\infty;ref} \frac{[L_{ref}]}{EC_{50;ref} + [L_{ref}]} = E_{\infty;test} \frac{[L_{test}]}{EC_{50;test} + [L_{test}]}$$

$$\frac{1}{[L_{test}]} = \frac{E_{\infty;test} EC_{50;ref}}{E_{\infty;ref} EC_{50;test}} \frac{1}{[L_{ref}]} + \frac{(E_{\infty;test} - E_{\infty;ref})}{E_{\infty;ref} EC_{50;test}} = a \frac{1}{[L_{ref}]} + b$$

Le facteur de biais est alors défini en fonction des coefficients a de deux voies différentes :

$$(4) \quad \beta_{2-1} = \log\left(\frac{a_1}{a_2}\right) = \log\left(\frac{E_{\infty;2;test} EC_{50;2;ref} E_{\infty;1;ref} EC_{50;1;test}}{E_{\infty;2;ref} EC_{50;2;test} E_{\infty;1;test} EC_{50;1;ref}}\right)$$

Cette définition du biais ligand-dépendant a pour avantage de pouvoir être estimée directement à partir des données brutes de régression sans même à avoir à connaître le K_D

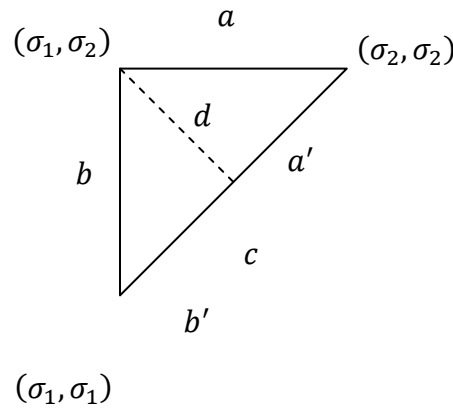
d'un ligand. Cependant, elle a pour désavantage d'être moins conservatrice que la définition du modèle opérationnel et produit donc davantage de faux positifs.

Le modèle opérationnel considère plutôt la relation entre le ligand étudié et un ligand équilibré via le coefficient de transduction (Rajagopal *et al*, 2011) :

$$(5.1) \quad \sigma = \log \left(\frac{\tau_{test}}{\tau_{ref}} \right)$$

Les paramètres σ pour deux voies de signalisation pour le ligand test sont alors comparés à ceux d'un ligand dit « équilibré » : la distance entre le point (σ_1, σ_2) et la droite définie par les points (σ_1, σ_1) et (σ_2, σ_2) est calculée.

Cette distance est reliée à l'altitude d'un triangle rectangle, qui donne la distance entre un point et une droite :



$$d^2 = a^2 - a'^2$$

$$\frac{a'}{a} = \frac{a}{c} \Rightarrow a' = \frac{a^2}{c}$$

$$\Rightarrow d^2 = \frac{a^2 b^2}{c^2} \Rightarrow d = \frac{ab}{c}, a > 0, b > 0, c > 0$$

Dans notre cas, on considère qu'un triangle formé au-dessus de la droite est un biais positif et l'inverse pour un triangle formé en dessous. Le signe du résultat est donc ici considéré :

$$d^2 = \frac{(\sigma_2 - \sigma_1)^2}{2} \Rightarrow d = \frac{|\sigma_2 - \sigma_1|}{\sqrt{2}}, \beta_{2-1} = \frac{\sigma_2 - \sigma_1}{\sqrt{2}} = \frac{1}{\sqrt{2}} (\Delta_{2-1} \Delta_{test-ref} \log \tau)$$

Comme le paramètre τ n'indique que l'efficacité d'un complexe ligand-récepteur à induire l'effet observé, une autre façon de représenter le biais pourrait aussi être de considérer conjointement τ et la capacité du ligand à induire la forme active du récepteur, soit :

$$(5.2) \quad \beta_{2-1} = \frac{1}{\sqrt{2}} \left(\Delta_{2-1} \Delta_{test-ref} \log \frac{\tau}{K_D} \right)$$

Considérant que le K_D d'un ligand ne change pas selon l'effet étudié (i.e. le système est les conditions expérimentales sont les mêmes), les expressions $\Delta_{2-1} \Delta_{test-ref} \log \tau$ et $\Delta_{2-1} \Delta_{test-ref} \log \frac{\tau}{K_D}$ sont équivalentes. Dans le cas contraire, l'erreur entre les deux expressions vaut (Kenakin et Christopoulos, 2013):

$$\Delta_{\frac{\tau}{K_D}} \beta_{2-1} = \log \left(\frac{K_{D;2;ref} K_{D;1;test}}{K_{D;2;test} K_{D;1;ref}} \right)$$

Évaluation du biais systémique associé à une mutation sur un récepteur

Dans la situation où le seul élément qui diffère d'un système à l'autre est le récepteur, l'analyse du biais est avant tout qualitative, sauf si le modèle opérationnel avec α est utilisé. Ceci dit, la comparaison des E_0 , E_∞ , EC_{50} et autres paramètres peut être utilisée pour dégager certaines tendances (Schönegge *et al*, 2017).

La transduction du signal par les GPCRs

Chronologie et enjeux

La transduction du signal est le phénomène qui transforme un signal extracellulaire, prenant habituellement la forme d'un ligand, en un signal intracellulaire, qui se manifeste souvent par une variation de la concentration d'une molécule intracellulaire appelée second messenger. La transduction du signal des GPCRs est un processus qui débute par la liaison de son ligand. Cette liaison provoque des changements conformationnels dans le GPCR, ce qui augmente son affinité pour ses protéines effectrices. La liaison du GPCR à ces protéines provoque à son tour leur activation et leur permet d'accomplir leurs fonctions.

La partie la plus importante et la plus mystérieuse de ce processus est la séquence d'événements entre la liaison du ligand et l'augmentation de l'affinité. Elle implique une série de résidus qui n'ont aucune interaction directe avec le ligand ou les effecteurs, mais qui sont néanmoins essentiels à l'activation des GPCRs. Les changements conformationnels de ces résidus restent encore à ce jour mal connus en raison de l'absence d'informations expérimentales sur la dynamique des GPCRs. Sans avoir d'information sur leur dynamique, les résidus importants sont néanmoins connus en raison des études de mutagenèse et de cristallographie effectuées sur une panoplie de GPCRs.

Études préalables de mutagenèse

De nombreuses études de mutagenèse ont été effectuées sur les GPCRs. Ces études sont variables en termes de complétude et de qualité des résultats, mais suffisantes pour déterminer quelles régions du GPCR sont plus impliquées dans la transduction du signal. Il ne semble exister que deux revues de littérature sur l'activation des GPCRs (Tehan *et al*, 2014) qui incluent des données de mutagenèse, mais un large éventail de publications sur de multiples GPCRs, incluant celles non-mentionnées dans ces revues, semblent confirmer la présence d'au moins trois régions importantes : une région appelée "Major H-Bond Network" (MHBN ou MHN), combinée au motif NPxxY (N7.49 - Y7.53), le motif DRY (D3.49 - Y3.51), aussi appelé "verrou 3-7", et une région de l'interface TM3-6 appelée "Hydrophobic Hindering Mechanism" (HHM) (Tehan *et al*, 2014; Trzakowski *et al*, 2012; Wescott *et al*, 2016; Katrich *et al*, 2013; Nygaard *et al*, 2009).

Tableau 3 : Proportion de mutations altérant le profil fonctionnel des GPCRs dans les régions susceptibles d'être impliquées dans l'activation des GPCRs.

Pour plus de détails, veuillez consulter l'Annexe III.

Mutation	Résidus	Total recensé	Total (%)
DRY	3.49/3.50/3.51	25	9,92%
MHBN et NPxxY	1.50/2.50/3.35/7.45/7.46/7.49/7.50/7.53	33	13,09%
HHM et WxP	3.40/3.43/6.40/6.41/6.44/6.48/6.50	75	29,76%
Autres		119	47,22%
Total		252	100,00%

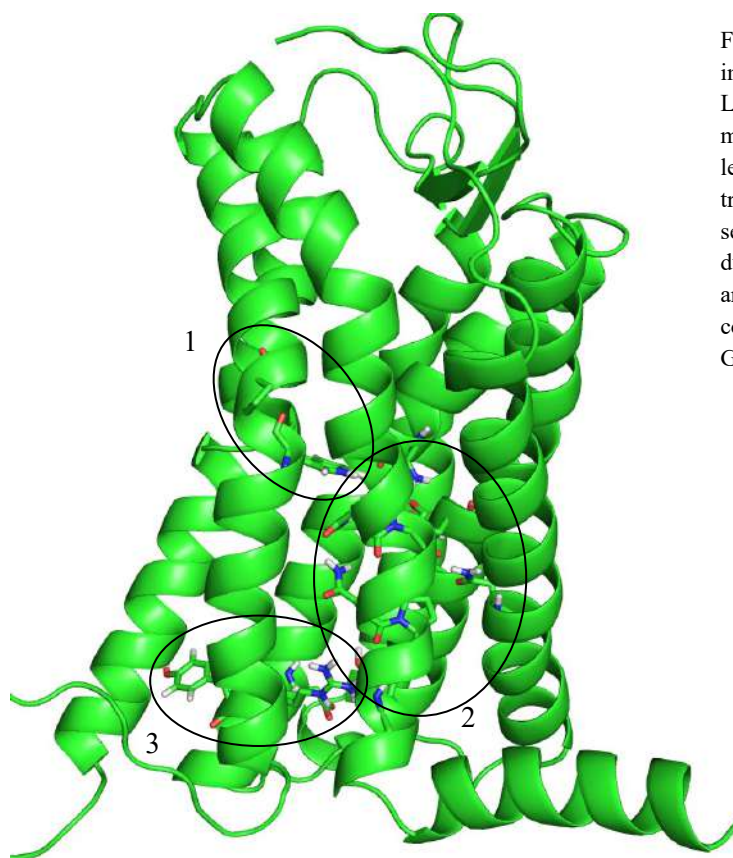


Figure7 : Régions des GPCRs rhodopsin-like impliquées dans la transduction du signal

Le plus près du site de liaison du ligand est le motif WxP (1). Suivent ensuite le MHBN et le motif NPxxY (2). Le motif DRY (3) se trouve plutôt éloigné des deux autres motifs et se trouve directement dans le site de liaison du GPCR avec les protéines G et les arrestines. Ces motifs sont entourés des contacts non-covalents conservés chez les GPCRs (Figure 3).

Toutes ces régions sont situées dans la section “charnière” des GPCRs : un groupe de résidus situés entre le site de liaison du ligand (extracellulaire) et le site de liaison des protéines effectrices du GPCR (intracellulaire). Ils contribueraient ainsi à moduler la réponse intracellulaire des GPCRs à la liaison d’un ligand sans directement être impliqués dans l’interaction avec l’un ou l’autre.

Le Major H-Bond Network et le motif NPxxY

Le “Major H-Bond Network” (MHBN ou MHN), ou en français le réseau majeur de ponts hydrogènes, est un ensemble de résidus regroupés spatialement autour du résidu D2.50 (N1.50, N7.49, occasionnellement 3.35 et 7.46) et dont les interactions sont régies par des ponts hydrogène. Plusieurs études ont fait état de l’importance de cet ensemble de résidus dans la signalisation des GPCRs : la mutation en alanine ou glycine des résidus N1.50 et N7.49 diminue fortement ou supprime l’activité du récepteur, de même que la mutation de D2.50 en asparagine (Marie *et al*, 1999; Baranski *et al*, 1999; Chen *et al*, 2001; Whistler *et al*, 2002; Bonde *et al*, 2010; Cabana *et al*, 2012; Cabana *et al*, 2015). Le résidu N3.35 a aussi été démontré important lors de l’activation du récepteur AT1 (Cabana *et al*, 2012; Cabana *et al*, 2015).

Le motif DRY

Le motif DRY est un des motifs les plus connus et est associé à la capacité du GPCR à s’activer et à lier les protéines G (Bonde *et al*, 2010). Les mutations dans le motif DRY au niveau de D3.49 vers plusieurs résidus tels que valine, asparagine, méthionine, histidine, etc., semblent toutes favoriser une forme active du récepteur, tandis que les mutations en alanine et en acide glutamique semblent provoquer l’effet inverse, voire prévenir la liaison d’un ligand au récepteur (Burger *et al*, 1999; Ballesteros *et al*, 2001; Li *et al*, 2001).

Certaines mutations, telle D3.49Y, provoquent des effets contradictoires sur des récepteurs différents (Li *et al*, 2001; Wang et Tao, 2011). Les mutations au niveau de type R3.50A semblent empêcher le récepteur de lier les protéines G de façon plus ou moins prononcée selon le récepteur, avec pour exception notable le récepteur A_{3A} Adénosine (Rhee *et al*, 2000; Chen *et al*, 2001; Feng et Song, 2003; Schneider *et al*, 2010; Bonde *et al*, 2010). Les

mutations de Y3.51 vers une phénylalanine ou une alanine semblent toutes diminuer l'affinité des GPCRs pour les protéines G (Rhee *et al*, 2000; Chen *et al*, 2001; Bonde *et al*, 2010).

Le Hydrophobic Hindering Mechanism et le motif WxP (tryptophane commutateur)

Le Hydrophobic Hindering Mechanism (HHM) et le motif WxP forment ensemble le lien entre le site de liaison du ligand et le site de liaison des protéines effectrices. En effet, W6.48 fait partie du site de liaison consensus des structures de GPCRs répertoriées entre 2000 et 2013. Ce tryptophane est aussi adjacent à un réseau de contacts conservé au-travers des structures de différents GPCRs (Venkatakrishnan *et al*, 2013). Ce réseau de contacts comprend des résidus adjacents dans l'espace ou inclus dans ce qui constitue le HHM.

Les résidus du HHM (3.40, 3.43, 6.40, 6.41 et 6.44) sont tous des résidus dont la mutation affecte de façon dramatique la capacité d'un GPCR à s'activer, ainsi que son affinité relative pour leurs protéines effectrices. Les résidus 3.43, 6.40 et 6.44 y sont particulièrement importants. Les résidus du HHM surgissent fréquemment dans l'étude de pathologies chez l'humain et dans des études de mutagenèse aléatoire chez la levure et ont donc été l'objet de (ou identifiés dans) plusieurs études (Spalding *et al*, 1998; Tonacchera *et al*, 1998; Baranski *et al*, 1999; Fanelli, 2000; Vaisse *et al*, 2000; Tao, 2000; Parnot *et al*, 2000; Tao *et al*, 2000; Greasley *et al*, 2002; Whistler *et al*, 2002; D'Antona *et al*, 2006; Bakker *et al*, 2008; Liu *et al*, 2007; Tao, 2008). Une étude de signalisation a été effectuée sur la L3.43 du FSHR et a pu démontrer que des mutations sur ce résidu sont en mesure de favoriser la synthèse d'AMPc ou l'internalisation de façons variables (Tao *et al*, 2000). Il s'agit cependant de la seule étude fonctionnelle effectuée sur des mutations de résidus du HHM. Les conclusions sont donc à valider sur d'autres GPCRs.

Dynamique des régions importantes dans l'activation des GPCRs

Bien que les résidus impliqués dans l'activation des GPCRs soient connus, le comportement de ces résidus à l'activation est mal connu. Les seules informations claires sont que le résidu R3.50 est retenu dans sa conformation inactive par des ponts-hydrogène et une interaction ionique au résidu D(E)3.49 ou un équivalent (e.g. le *ionic lock* avec

D(E)6.30 ou un pont-hydrogène avec Y7.53). Lors de l'activation du GPCR, cette interaction est brisée et R3.50 peut se positionner de façon à interagir avec les protéines G (Tehan *et al*, 2014; Trzakowski *et al*, 2012; Katrich *et al*, 2013; Nygaard *et al*, 2009; Elgeti *et al*, 2013; Mahalingam *et al*, 2008).

La dynamique des autres régions importante dans l'activation des GPCRs n'est pas claire si uniquement le mouvement des résidus est considéré. Cependant, si on considère la dynamique à l'échelle des structures secondaires et de la structure tertiaire, il y a deux événements dont l'occurrence est avérée. Le premier est la distorsion de l'hélice du TM6 entre P6.50 et F6.44, ce qui provoque le déplacement de F6.44 vers le TM5 et la rotation de la portion intracellulaire du TM6. Ce mouvement est observé dans la majorité des structures de récepteurs cristallisés avec des agonistes en l'absence de protéines effectrices (voir Tableau 1) et permettrait au TM6 d'effectuer le changement conformationnel majeur qui est associé à la liaison des protéines G (Tehan *et al*, 2014; Trzakowski *et al*, 2012; Nygaard *et al*, 2009; Dror *et al*, 2011; Rasmussen *et al*, 2011; Carpenter *et al*, 2016; Liang *et al*, 2017; Zhang *et al*, 2017; Liang *et al*, 2018; Koehl *et al*, 2018; Draper-Joyce *et al*, 2018; Garcia-Nafria *et al*, 2018). Le second est l'ouverture d'un pore entre les TM2, 3, 6 et 7, qui permet l'infiltration d'eau au cœur de la structure du GPCR et, par conséquent, sa dilatation de l'ouverture de sa portion intracellulaire à la liaison de ses effecteurs. Ce pore est bordé par le HHM d'un côté et par le MHBN de l'autre. Il est aussi suspecté que les molécules d'eau complètent le MHBN et permettent la formation de ponts-hydrogène essentiels à l'activation des GPCRs (Tehan *et al*, 2014; Cabana *et al*, 2012; Nygaard *et al*, 2009). Pour ce qui est des changements conformationnels menant à une interaction avec les arrestines, le seul récepteur qui a jusqu'à maintenant été cristallisé avec une arrestine est la rhodopsine (Kang *et al*, 2015; Latorraca *et al*, 2018; Zhou *et al*, 2016). Ce qui peut y être observé est la formation d'une interaction entre le TM6 et le TM7 au niveau intracellulaire, via le résidu 6.40 et le résidu 7.53. Le système présente un réarrangement au niveau du TM6 moins prononcé et un du TM7 similaire à ceux associés à l'interaction avec les protéines G.

Interaction des GPCRs avec leurs partenaires de liaison

Le mode de liaison des GPCRs aux hétérotrimères Gs et Gi ont récemment été déterminés pour différents récepteurs (voir Tableau 1), ce qui permet de comparer leur mode de liaison avec plus de fiabilité. Le site de liaison des protéines G implique principalement les TM3, 5 et 6 et les ICL2 et 3 (Rasmussen *et al*, 2011; Carpenter *et al*, 2016; Liang *et al*, 2017; Zhang *et al*, 2017; Liang *et al*, 2018; Koehl *et al*, 2018; Draper-Joyce *et al*, 2018; García-Nafria *et al*, 2018; Wong *et al*, 2003), tandis que l'arrestine visuelle se lie à la rhodopsine par l'intermédiaire des TM2, 3, 5, 6 et la région entre TM7 et H8, ainsi que des ICL 2 et 3 (Kang *et al*, 2015; Zhou *et al*, 2016; Zhou *et al*, 2017). Dans les deux cas, la majorité des interactions entre le GPCR et la protéine G ou l'arrestine sont des interactions hydrophobes. Plus récemment, une étude, la première sur la structure des complexes GPCR-GRK, sur l'estimation de la conformation du complexe β_2 -AR-GRK5 a été publiée. Elle indique que l'interface entre les deux protéines implique l'ICL2, l'ICL3 et l'H8-C-term du β_2 -AR et semble s'infiltrer dans la pochette de liaison cytoplasmique du GPCR de la même façon que les protéines G, mais sans provoquer des changements conformationnels aussi importants (Komolov *et al*, 2017).

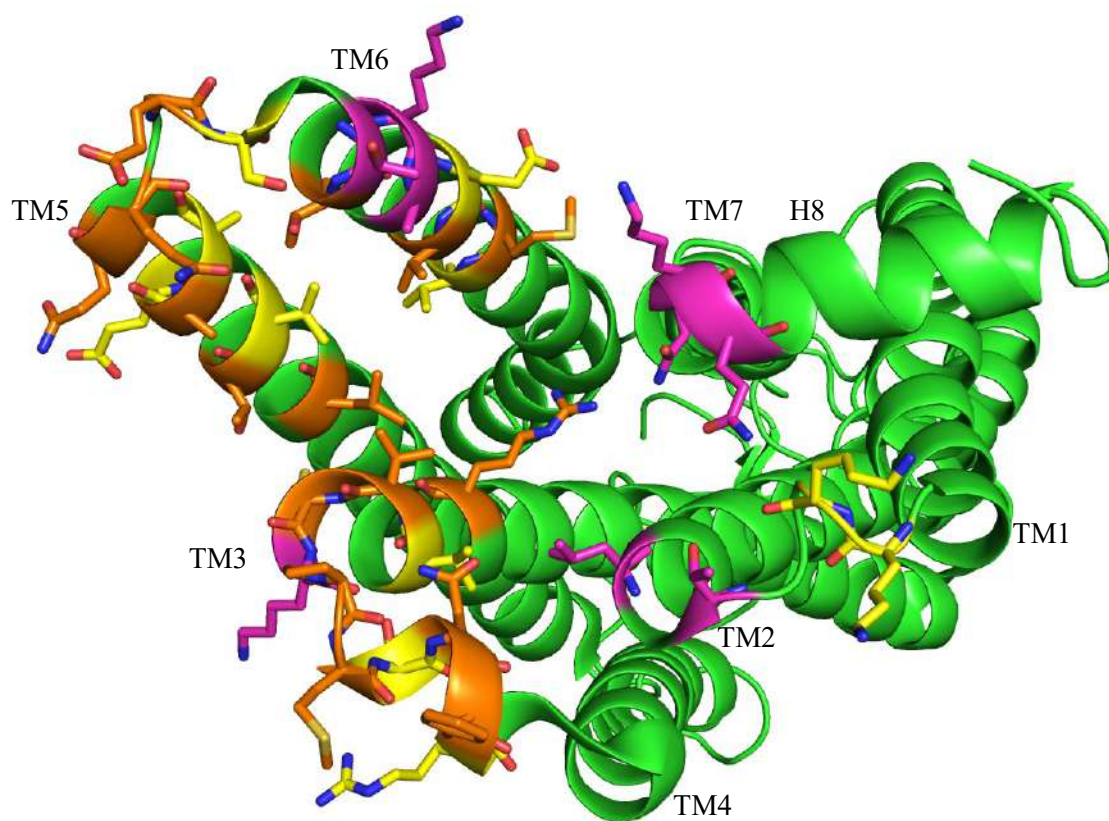


Figure 8 : Interfaces d'interaction entre les GPCRs et les protéines G et les arrestines

Les résidus à proximité de l'interface entre les GPCRs et les protéines G sont marqués en jaune, ceux de l'interface avec les arrestines, en magenta, et les résidus communs, en orange. L'interaction avec les protéines G semble se concentrer au niveau des TM3, 5 et 6, tandis que l'interaction avec les arrestines semble s'étendre également aux TM2 et 7.

Malgré la présence d'informations sur la nature des complexes GPCR-protéine G, GPCR-arrestine et GPCR-GRK, l'importance de ces informations reste discutable. En effet, une étude de single-molecule fluorescence RET (smFRET) a démontré qu'en présence d'un agoniste, de GDP et de GTP, la population du complexe GPCR-protéine G tel qu'observé dans les structures de cristaux ou de Cryo-EM devient beaucoup plus petite. De plus, le processus d'activation des protéines G impliquerait des complexes GPCR-protéine G dont le TM6 serait dans une conformation plus proche des structures dans un état inactif ou lié à un agoniste sans protéine G ou anticorps (Gregorio *et al*, 2017). Une autre étude a aussi été publiée sur la dynamique des événements pré- et post-complexe et semble indiquer que les événements pré-complexe impliqueraient l'ICL1 et l'H8 des GPCRs (Sounier *et al*, 2015).

Dans le cas des arrestines, les seules structures de complexes existantes sont celles de la rhodopsine avec une arrestine visuelle, ce qui rend difficile d'extrapoler sur la structure de récepteurs rhodopsine-like en complexe avec des arrestines non-visuelles.

Structure et signalisation biaisée

Peu d'information existe actuellement sur la relation entre structure et signalisation biaisée. Une étude de RMN ^{19}F sur le récepteur β_2 -adrénergique indique que la signalisation biaisée pro-protéine G implique des changements conformationnels dans la région intracellulaire des TM6 et 7, tandis qu'une signalisation biaisée pro-arrestine implique des changements conformationnels dans la région intracellulaire du TM7 seulement (Liu *et al*, 2012). Une étude de cristallographie a comparé les récepteurs 5-HT_{1B} et 5-HT_{2B} en complexe avec l'ergotamine, un ligand présentant un biais pro-arrestine important chez 5-HT_{2B}, mais pas chez 5-HT_{1B}. Dans la structure de 5-HT_{1B}, le résidu 6.44 est dans une position active et le TM6 subit une légère rotation, ce qui tend à indiquer que le récepteur se trouve dans une région de l'espace conformationnel propice à la liaison des protéines G. Ce changement n'est pas observé chez 5-HT_{2B}. Un changement commun observé est le réarrangement de la portion intracellulaire du TM7, qui semble plus prononcé chez 5-HT_{2B} (Wacker *et al*, 2013). Un article sur la découverte d'un ligand biaisé du récepteur μ -opioïde fait état d'une interaction plus forte du ligand biaisé pro-protéine G et du résidu 3.32, mais sans produire de structure expérimentale (Manglik *et al*, 2016). Un dernier article, qui utilise la dynamique moléculaire, parle plutôt d'une interaction entre le récepteur D2R et un ligand non-biaisé au niveau du résidu 5.42 qui serait brisée et remplacée par une interaction avec l'ECL2 dans le cas d'un ligand biaisé pro-arrestine (McCorvy *et al*, 2018). Mises ensemble, ces informations indiquent que les interactions entre les ligands et les TM3 et 5 semblent importants pour provoquer une signalisation biaisée et que la signalisation via les protéines G implique davantage des changements conformationnels au niveau du TM6, tandis que la signalisation via les arrestines implique plutôt des réarrangements au niveau du TM7. Le résidu 6.44 semble aussi devoir être en position « active-like » pour permettre la signalisation via les protéines G. Il faut cependant rester vigilant, puisque les complexes rhodopsine-arrestine ont aussi le résidu 6.44 dans une conformation « active-like »,

contrairement aux structures de rhodopsine inactive. De plus, ces informations sont tirées de quatre récepteurs différents, et ce dans des contextes totalement différents.

Le récepteur de l'apeline

Profil de signalisation, fonctions principales et relation entre signalisation et effets physiologiques

Le récepteur de l'apeline (APLNR, ou APJ) est un GPCR aux multiples rôles et applications. APJ possède deux ligands endogènes : l'apeline, avec les isoformes principales apeline-36, apeline-17, apeline-13 et [Pyr¹]apeline-13 (Medhurst *et al*, 2003), et ELABELA, avec pour isoformes ELABELA-32, ELABELA-21 et ELABELA-11 (Murza *et al*, 2019).

Il est exprimé dans plusieurs tissus et organes dans le corps, notamment l'endothélium vasculaire, le cœur, les reins et le système nerveux central (O'Carroll *et al*, 2013). Ses fonctions principales sont associées au système cardiovasculaire et à la fonction rénale. APJ en complexe avec l'apeline provoque la vasodilatation et présente des effets inotropes positifs au niveau du cœur, un effet cardioprotecteur en conditions de stress cardiaque, ainsi qu'un effet antidiurétique (O'Carroll *et al*, 2013). Ces effets physiologiques associés à APJ en font une cible intéressante pour deux applications importantes : l'insuffisance cardiaque, principalement l'hypertension artérielle pulmonaire (Kim, 2014), et le choc septique (Chagnon *et al*, 2017; Coquerel *et al*, 2018).

APJ est un récepteur au profil de signalisation relativement simple. Lorsqu'il se trouve en complexe avec l'apeline-13, il signale principalement via $G\alpha_{i1}$, $G\alpha_{i2}$ et $G\alpha_o$ (Masri *et al*, 2006; Besserer-Offroy *et al*, 2018; Murza *et al*, 2012; Murza *et al*, 2015) et peut lier autant la β -arrestine 1 que la β -arrestine 2 (Besserer-Offroy *et al*, 2018; Murza *et al*, 2012; Murza *et al*, 2015). Certaines preuves expérimentales indirectes indiquent aussi une potentielle signalisation via G_q (Szokodi *et al*, 2002) ou G_{13} (Kang *et al*, 2013), mais ces informations ne semblent pas avoir été explorées davantage par la littérature plus récente.

La relation entre la signalisation d'APJ et ses effets physiologiques n'a pas encore été démontrée de manière fiable. En effet, une seule étude a jusqu'à maintenant établi un lien

entre le biais de signalisation de systèmes ligand-récepteur et un effet d'APJ : le biais pro-arrestine ou anti-arrestine par rapport aux protéines G semble corrélé avec un effet hypotenseur chez le rat (Besserer-Offroy *et al*, 2018).

Structure d'APJ et relation structure-activité de ses ligands

La relation structure-activité d'APJ en complexe avec ses ligands a été l'objet de multiples études (Besserer-Offroy *et al*, 2018; Murza *et al*, 2012; Murza *et al*, 2015; Murza *et al*, 2019; El Messari *et al*, 2004; Lee *et al*, 2005; Ceraudo *et al*, 2014; Margathe *et al*, 2014; Brame *et al*, 2015; Murza *et al*, 2016; Murza, Sainsily *et al*, 2017; Tràn *et al*, 2018; Fan *et al*, 2003; Gerbier *et al*, 2015). De ces études, il se dégage sur le plan fonctionnel que le résidu phénylalanine en position C-terminale dans l'apeline-13 et son équivalent sur ELABELA semblent importants pour la fonction du récepteur, particulièrement en ce qui a trait à son internalisation. De plus, au moins un résidu ou groupement fonctionnel fortement basique semble essentiel pour la liaison au récepteur.

Dans la structure publiée du récepteur en complexe avec un agoniste peptidique macrocyclique (Ma *et al*, 2017), l'agoniste utilisé ressemble fortement à l'apeline-17 et fournit par conséquent des informations sur le mode de liaison des différents isoformes de l'apeline. En prenant l'apeline-17 comme repère pour la numérotation, le résidu F17 interagit principalement avec des résidus des TM1, 2 et 7 et les R6 à L9 forment localement un feuillet beta antiparallèle avec la queue N-terminale d'APJ. Ceci dit, comme il s'agit d'une structure issue de cristallographie aux rayons X, la structure est statique et n'indique rien sur les multiples possibles modes de liaison des isoformes de l'apeline, et encore moins sur ceux des isoformes d'ELABELA. De plus, la cavité où se lie la portion C-terminale du ligand macrocyclique forme un immense « bol » hydrophobe, ce qui laisse penser que les résidus hydrophobes en position C-terminale peuvent adopter différentes conformations dans cette cavité.

Présentation du projet

Déterminants structuraux de la signalisation biaisée d'APJ : un cycle de l'ignorance

Le peu d'information qui existe sur la relation entre structure, dynamique et signalisation biaisée rend le processus de design de ligands ou de mutants biaisés plus semblable à une loterie qu'à un véritable processus de design. En effet, les résidus ou régions à cibler dans la pochette de liaison des ligands, ou dans les régions impliquées dans la transduction du signal, afin de provoquer une signalisation biaisée sont en général inconnus, si résidus ou régions communes il y a. De plus, ce problème fait partie d'un problème plus large et beaucoup plus intéressant d'un point de vue fondamental, mais tout aussi peu compris : comment fonctionne la signalisation biaisée au niveau de la structure des GPCRs? Dans les deux cas, il y a peu d'éléments de réponse, ce qui rend le processus de design de ligands ou de mutants biaisés ardu et peu fiable.

Une quantité d'information importante manque encore au bagage existant avant que qui que ce soit puisse tirer des conclusions sur comment le processus de signalisation biaisée se produit sur le plan structural. Pour produire cette information, il existe deux approches : l'approche directe et l'approche indirecte. L'approche directe consiste en l'obtention de la structure de complexes ligand-GPCR directement via cristallographie aux rayons X ou cryo-EM. L'information structurale apportée par cette approche est beaucoup plus crédible que l'approche indirecte, mais manque d'information sur la dynamique du complexe observé. De plus, l'obtention d'un cristal demande beaucoup de temps et de ressources et n'est jamais garantie. L'approche indirecte consiste en l'analyse de la dynamique d'un complexe ligand-GPCR, via la RMN, la dynamique moléculaire ou le smFRET. L'information structurale apportée ici est beaucoup plus exhaustive et inclut la dynamique des complexes. Cependant, comme les techniques existantes de détermination de structures complètes utilisant la RMN sont restreintes aux petites molécules, l'information obtenue par RMN est incomplète et se limite au décompte des différents états conformationnels de régions limitées du GPCR. De plus, même dans le cas d'une analyse incomplète de la structure du GPCR, la présence de sondes réactives, principalement à base de ^{13}C ou de ^{19}F en RMN, sont requises et les observations sont souvent limitées aux méthionines et aux autres résidus nucléophiles, ce qui demande un travail de mutagenèse du GPCR d'ampleur

variable afin de limiter les signaux indésirables (Ye *et al*, 2018; Liu *et al*, 2012; Sounier *et al*, 2015; Kofuku *et al*, 2012; Nygaard *et al*, 2013; Kim *et al*, 2013; Ye *et al*, 2016). Pour ce qui est de l'information apportée par les simulations de dynamique moléculaire, elle est certes complète et facile à obtenir, mais reste en essence théorique et requiert une validation expérimentale subséquente, en plus d'être limitée à des échelles de temps très courtes d'un point de vue biologique.

Les techniques d'analyse structurale existent, mais pour obtenir de l'information sur l'aspect structural de la signalisation biaisée chez un GPCR, il faut des ligands, ou encore des mutants biaisés du récepteur. Ceci dit, il est présentement impossible de concevoir des ligands ou des mutants biaisés de façon fiable pour la majorité des GPCRs, ce qui inclut APJ. Le développement des connaissances sur l'aspect structural de la signalisation biaisée des GPCRs est donc coincé dans un cycle de l'ignorance : pour obtenir des informations, il faut des ligands ou des mutants, et pour concevoir de tels ligands ou mutants, il faut des informations. La seule façon de se sortir de ce cycle est de continuer d'ajouter des informations, trouvées le plus souvent par chance, au bagage existant. L'intérêt d'APJ dans ce contexte est alors important : il s'agit d'un récepteur dont la structure et la fonction sont relativement peu étudiés, ce qui augmente d'autant la valeur des informations qu'on pourrait en tirer. De plus, comme il est cristallisé avec un agoniste semblable à l'apeline-17, l'analyse du système APJ-apeline par dynamique moléculaire risque moins d'être faussée par la mauvaise position du ligand ou par une conformation initiale trop loin de ses formes actives.

La question qui définit le présent projet est donc : « Quels sont les déterminants structuraux de la signalisation biaisée du récepteur de l'apeline? » L'hypothèse de travail est que la réponse se trouve dans les mêmes régions que celles impliquées dans la transduction du signal : les endroits sont les mêmes, mais les changements sont différents.

Approche et objectifs

L'approche utilisée pour explorer cette hypothèse comprend six étapes :

1. Échantillonner l'espace conformationnel d'APJ en complexe avec l'apeline (système WT-apeline) à l'aide de simulations de dynamique moléculaire, avec APJ seul (système WT-apo) comme contrôle.
2. Identifier les résidus de la région transmembranaire impliqués dans des changements conformationnels lents dans le système WT-apeline n'apparaissant pas dans le système WT apo.
3. Concevoir et produire des mutants aux positions à proximité des résidus les plus importants en s'inspirant de la littérature.
4. Analyser la fonction de ces mutants et identifier ceux présentant un profil fonctionnel intéressant, c'est-à-dire :
 - Liaison à l'apeline-13
 - Expression sur la membrane cytoplasmique
 - Un des deux profils suivants :
 - Activité constitutive ou maximale non-nulle par rapport à l'activité basale d'APJ WT, mais exclusivement pour $G\alpha_i$ ($G\alpha_{i1}$ est ici utilisée)
 - Activité constitutive ou maximale non-nulle par rapport à l'activité basale d'APJ WT, mais exclusivement pour les β -arrestines (β -arrestine-2 est ici utilisée)
5. Échantillonner l'espace conformationnel des mutants d'APJ dans les conditions les plus intéressantes et diversifiées (i.e. type de biais, en complexe avec ou sans l'apeline) à l'aide de simulations de dynamique moléculaire.
6. Identifier les résidus de la région transmembranaire impliqués dans des changements conformationnels lents et comparer leurs changements conformationnels avec ceux observés uniquement chez APJ WT avec apeline.

L'emphasis est mise sur les changements conformationnels lents car il s'agit de ceux dont l'importance est la plus grande d'un point de vue biologique (Noé et Nüske, 2013; Pérez-Hernandez *et al*, 2013; Schwantes et Pande, 2013). Par exemple, la dynamique chaotique et

rapide d'une chaîne latérale est de peu d'intérêt par rapport à un changement conformationnel concerté de plusieurs résidus et il advient que ces changements sont habituellement beaucoup plus lents. Les protocoles de biologie expérimentale ont déjà été optimisés pour APJ et seront un peu modifiés tel que détaillé dans la section Matériel et Méthodes, mais les différentes étapes de biochimie structurale théorique présentent leurs propres défis dont il sera question dans les prochaines sections.

Matériel et méthodes

Présentation des méthodes théoriques utilisées

La dynamique moléculaire : théorie et pratique

La dynamique moléculaire est une méthode de chimie computationnelle qui permet l'étude directe, quoique théorique, de systèmes normalement impossibles à observer directement à l'échelle microscopique dans des conditions normales. Cette méthode utilise les équations du mouvement de la physique classique, avec m la masse, \mathbf{r} , \mathbf{v} et \mathbf{a} la position, la vitesse et l'accélération vectorielles, respectivement, et V et \mathbf{F} l'énergie potentielle et la force correspondante, respectivement, pour estimer le comportement du système étudié :

$$\begin{aligned}\mathbf{r} &= \mathbf{r}_0 + \mathbf{v}t \\ \mathbf{v} &= \mathbf{v}_0 + \mathbf{a}t \\ \mathbf{F} &= m\mathbf{a} = -\nabla V(\mathbf{r})\end{aligned}$$

Les variables vectorielles (vecteurs colonne) sont ici représentées en gras. L'indice zéro représente la grandeur au temps zéro. La grandeur $-\nabla V(\mathbf{r})$ est dictée par un ensemble de paramètres et d'équations appelé « champ de force », qui dicte le comportement de chaque atome en fonction de sa relation et de sa proximité avec les autres. Dans ce projet, le champ de force GROMOS 54a7 est utilisé, puisqu'il permet une définition du système en mode *united atom*, est optimisé spécifiquement pour l'étude de macromolécules biologiques, avec une paramétrisation simple des molécules d'eau, et a été corrigé par rapport à la version précédente pour bien stabiliser les hélices alpha, ce qui devrait permettre une représentation assez fiable d'un GPCR, en plus d'accélérer grandement l'échantillonnage de son espace conformationnel (Abraham *et al* , 2018; Schmid *et al*, 2011).

À cela s'ajoute une méthode de contrôle de la température, appelée « Thermostat », et une méthode de contrôle de la pression, appelée « Barostat ». Le thermostat accomplit sa fonction en contrôlant l'énergie cinétique moyenne K du système en fonction de la température T voulue (Abraham *et al* , 2018) :

$$K = \frac{1}{2} \sum_{i=1}^N m_i v_i^2$$

$$K = \frac{3}{2} k_B T (N - 1)$$

N représente le nombre total d'atomes. Ici, comme la condition périodique aux limites est utilisée et que le système est exempt d'autres contraintes, le nombre de degrés de liberté rotationnels tombe à zéro et il ne reste que les degrés de liberté translationnels (Abraham *et al* , 2018). Dans notre cas, une modification de l'équation de base de la dynamique moléculaire, l'équation de Langevin, sert de thermostat :

$$(6) \quad m\mathbf{a} = \mathbf{F} - \gamma m\mathbf{v} + \mathbf{R}$$

\mathbf{R} représente ici un bruit blanc gaussien (utilisé ici comme force aléatoire) de moyenne nulle et de variance égale à $2m\gamma k_B T$ dans chaque dimension. $-\gamma m\mathbf{v}$ est une force dissipative visant à limiter l'augmentation de l'énergie cinétique du système provoquée par \mathbf{R} . La constante γ est fixée à $0,5 \text{ ps}^{-1}$ car cette valeur donne des forces dissipatives inférieures à la friction interne de l'eau, son inverse (63420 ans) est bien plus grand que les autres échelles de temps du système et sa valeur est suffisamment élevée pour constituer un thermostat efficace (Abraham *et al*, 2018). Comme les autres thermostats, l'équation de Langevin fonctionne en redimensionnant la vitesse des molécules de façon à ce que leur ensemble corresponde à une distribution de Maxwell-Boltzmann à la température correspondante (Abraham *et al*, 2018; Hünenberger, 2005). Cette équation présente aussi d'autres avantages qui seront abordés plus loin.

Le barostat, quant à lui, accomplit sa fonction en redimensionnant le système dans la simulation. Pour ce faire, la pression P est calculée à partir de la forme matricielle du théorème du viriel. Dans le cas qui nous intéresse, le système est semi-isotrope, c'est-à-dire isotrope dans les directions parallèles à la membrane (x et y) et isotrope, mais séparé, dans la direction perpendiculaire :

$$\begin{aligned}
\mathcal{K} &= \frac{1}{2} \sum_{i=1}^N m_i \mathbf{v}_i \mathbf{v}_i^T \\
\Xi &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=i+1}^N m_i \mathbf{r}_{ij} \mathbf{F}_{ij}^T \\
P &= \frac{2}{V} (\mathcal{K} - \Xi) \\
P_{xy} &= \frac{1}{2} (P_{xx} + P_{yy}), P_z = P_{zz}
\end{aligned}$$

L'exposant T représente la transposée. Le barostat utilisé ici est celui de Parrinello-Rahman. Il fonctionne en redimensionnant la boîte qui délimite le système durant la simulation :

$$\begin{aligned}
\frac{d^2}{dt^2} b_x &= b_x b_y b_z W_{xy} \left(\frac{d}{dt} b_x \right)^{-1} (P_{xy} - P_{ref}) \\
\frac{d^2}{dt^2} b_y &= b_x b_y b_z W_{xy} \left(\frac{d}{dt} b_y \right)^{-1} (P_{xy} - P_{ref}) \\
\frac{d^2}{dt^2} b_z &= b_x b_y b_z W_z \left(\frac{d}{dt} b_z \right)^{-1} (P_z - P_{ref})
\end{aligned}$$

b_x représente ici la dimension en x de la boîte, de même pour b_y et b_z . Les paramètres W représentent la force du couplage. Cet ajustement modifie la position des atomes en conséquence, de façon à ce que tous les atomes entrent dans la boîte du système (Abraham *et al*, 2018; Nosé et Klein, 1983).

L'ensemble de ces étapes est répété à chaque itération de la simulation, soit dans notre cas chaque 5 fs. Ce pas d'intégration substantiel accélère considérablement les simulations et n'est possible que par la formation de sites virtuels, soit la création d'un seul atome représentant à la fois l'atome lourd et les hydrogènes qui y sont liés, avec l'ajout de contraintes appropriées pour simuler un comportement normal de la molécule, et l'utilisation du résolveur de contraintes LINCS (Abraham *et al*, 2015).

La conformation du système est sauvegardée moins fréquemment, par exemple chaque 50 ps, mais toujours à la même fréquence, et forme une suite de conformations appelée « trajectoire » :

$$\mathcal{T} = \{(\mathbf{r}_1, \mathbf{v}_1), \dots, (\mathbf{r}_N, \mathbf{v}_N)\}$$

L'ensemble des trajectoires peut alors être utilisé à des fins d'analyse statistique.

Limitations intrinsèques, limitations techniques et comment y remédier

Malgré le fait que la dynamique moléculaire apporte des informations de bien plus grande précision et complétude que les techniques expérimentales d'analyse structurale des GPCRs, quelques limitations importantes sont à considérer :

- L'exactitude de ces informations reste discutable en raison des approximations et des systèmes utilisés pour l'optimisation des champs de force.
- Les conditions périodiques aux limites donnent lieu à certaines problématiques qui peuvent mener à la présence d'artefacts dans les résultats.
- L'ergodicité de l'ensemble des simulations n'est pas garanti, ce qui pose problème lors de l'estimation des différentes grandeurs thermodynamiques du système.

Chacun de ces problèmes représente une limitation intrinsèque ou technique associée à la dynamique moléculaire. Des solutions existent pour y remédier, mais peu sont réalisables dans l'état actuel des capacités des ordinateurs. Par exemple, les limitations associées aux approximations requerraient l'utilisation des équations du mouvement de la physique quantique, ce qui est irréalisable avec les capacités computationnelles actuelles des ordinateurs : la complexité des algorithmes existants est simplement trop grande.

Les deux autres limitations présentent cependant quelques pistes de solutions. Dans le cas des conditions périodiques aux limites, le problème est que le système, bien que délimité par une boîte, est répété en copie exacte à l'infini dans chaque direction (Abraham *et al*, 2018), ce qui donne un caractère cristallin au système. Ce type de contrainte ajoute diverses sources d'erreurs, la plus importante étant les artefacts d'arrangement cristallin associés aux interactions anormales d'une molécule avec elle-même, comme par exemple

l'interaction de la portion extracellulaire d'un GPCR avec sa portion cytoplasmique. Pour remédier à ce problème, l'ajustement de la taille du système est la seule solution accessible. Ceci dit, une attention particulière doit être apportée à ne pas trop augmenter la quantité d'atomes, car dans un champ de force, il existe une certaine quantité d'interaction impliquant deux atomes à proximité dans l'espace. La quantité de ces interactions augmente de façon quadratique avec nombre d'atomes.

En effet, si on ne compte qu'une seule fois les interactions entre deux atomes, le nombre de forces à calculer vaut :

$$N_F = \sum_{i=1}^N \sum_{j=i+1}^N 1 = \sum_{i=1}^N (N - i) = \frac{N(N - 1)}{2} \in O(n^2)$$

Ainsi, un système trop gros limitera l'importance des artefacts d'arrangement cristallin autant que la rapidité de calcul des trajectoires. Comme le rayon seuil (à partir duquel la force calculée devient négligeable) pour les interactions électrostatiques se situe à 12 Å pour les simulations effectuées ici, la somme de la distance entre l'atome du récepteur le plus près de la limite inférieure en z de la boîte et ladite limite et de la distance correspondante pour la limite supérieure en z doit au minimum égaler cette valeur seuil. Dans notre cas, cette distance avoisine les 18 Å, ce qui limite les artefacts associés aux conditions périodiques aux limites sans trop augmenter la taille du système, en plus de laisser une marge de manœuvre pour le redimensionnement du système par l'algorithme du barostat. Les interactions de plus longue distance sont traitées avec l'approximation du Particle Mesh Ewald (Abraham *et al*, 2018).

L'autre limitation majeure associée à l'estimation de paramètres thermodynamiques, comme les populations de différents états ou la cinétique de transition entre différents états d'un système, est le fait que, en particulier pour les systèmes biologiques de grande taille, l'ergodicité des trajectoires de dynamique moléculaire est loin d'être garantie (Coveney et Wan, 2016). Or, une propriété thermodynamique d'un ensemble à l'équilibre et la moyenne

de cette même propriété obtenue d'une trajectoire continue dans le temps ne sont égales que si la durée de la trajectoire tend vers l'infini.

C'est-à-dire, que, pour une propriété G :

$$\lim_{t \rightarrow \infty} \left(\bar{G} - \frac{1}{t} \int_0^t G(t) dt \right) = 0$$

$$\lim_{t \rightarrow \infty} \left(\bar{G} - \frac{1}{N} \sum_{(\mathbf{r}_i, \mathbf{v}_i) \in \mathcal{T}} G(\mathbf{r}_i, \mathbf{v}_i) \right) = 0$$

Or, les trajectoires générées en dynamique moléculaire sont loin d'être de durée infinie, et souvent ne sont même pas de durée suffisante pour observer les changements conformationnels les plus rares. Cela vient nécessairement fausser les estimations des propriétés thermodynamiques du système issues de ces trajectoires.

Deux solutions à ce problème se présentent. La première solution est d'utiliser la sensibilité des simulations de dynamique moléculaire aux conditions initiales pour accélérer la convergence du système vers la vraie distribution à l'équilibre. En effet, un échantillonnage avec une grande variété de conditions initiales permet l'exploration plus rapide de l'ensemble des conformations et vitesses accessibles (l'espace des phases) qu'une seule longue trajectoire. Cette propriété s'exprime mathématiquement par le concept de mélange, c'est-à-dire que, dans le cas où un ensemble de conditions initiales différents est choisi au lieu d'une seule, l'ensemble des systèmes issus ces conditions (A) se dispersera à chaque itération de l'algorithme de dynamique moléculaire (T) de façon à en venir à occuper n'importe quel sous-ensemble de l'espace des phases du système (B) :

$$\lim_{n \rightarrow \infty} P(T^n(A)|B) = P(A)$$

Dans le cas d'une seule simulation prolongée, une telle dispersion est impossible et son occupation d'un sous-ensemble ne converge qu'en moyenne, avec des instants où la simulation se trouve dans le sous-ensemble et d'autres non :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} P(T^n(A)|B) = P(A)$$

De façon plus imagée, la comparaison est similaire entre la dispersion passive d’une goutte de colorant dans un verre d’eau par rapport à celle d’une bille de métal de taille équivalente. Le colorant (simulations multiples) se disperse rapidement et en vient à occuper la totalité du verre, tandis que la bille (une simulation) reste entière, et ce même si on agite vigoureusement le verre, et n’occupera donc jamais la totalité du verre en même temps. De plus, il ne faut que quelques secondes au colorant pour se répandre partout dans le verre, tandis que la bille prendra plus de temps à visiter chaque endroit. Ces conclusions s’appliquent essentiellement à des ensembles continus de l’espace des phases, mais une conclusion similaire a été tirée à partir d’un nombre fini de conditions de départ et de simulations (Coveney et Wan, 2016). Dans notre cas, des ensembles de 8 simulations de 200 ns chacune, pour un total de 1,6 μ s, sont effectuées à chaque itération du protocole de génération de trajectoires. La nature aléatoire de la dynamique de Langevin devrait aider d’autant l’échantillonnage en accélérant la dispersion des différents systèmes.

Une autre problématique plus spécifique aux systèmes biologiques est celle du temps de transition entre deux états conformationnels du système. En effet, la cinétique de transition de plusieurs changements conformationnels majeurs et de haute importance d’un point de vue fonctionnel sont de l’ordre de la microseconde ou de la milliseconde, soit des échelles de temps encore difficilement accessibles en dynamique moléculaire des systèmes biologiques. Cela signifie que, considérant que les trajectoires générées sont de longueur semblable à l’échelle de temps de ces transitions, la probabilité d’observer une telle transition est faible. Ainsi, si toutes les trajectoires utilisées pour l’échantillonnage débutent avec une conformation donnée, mais des vitesses aléatoires, l’échantillonnage des conformations au-delà de cette transition rare par rapport à la conformation de départ sera sévèrement limité. Pour remédier à ce problème, la solution consiste à non seulement faire varier les vitesses de départ, mais aussi les conformations. En utilisant comme conformations de départ les régions de l’espace des configurations les moins explorées, il

est possible d'accélérer grandement la convergence de l'échantillon des différentes trajectoires vers l'ensemble thermodynamique du système (Weber et Pande, 2011).

Ainsi, la solution proposée pour un échantillonnage optimal de l'espace conformationnel d'un système biologique est le suivant :

1. Effectuer 8 simulations en parallèle du système avec la même conformation de départ, mais des vitesses différentes
2. Identifier les états conformationnels les moins échantillonnés
3. Effectuer 8 autres simulations avec les états identifiés en 2 et en réinitialisant les vitesses des atomes de façon aléatoire et conforme à la distribution de Maxwell-Boltzmann.
4. Répéter 2 et 3 jusqu'à convergence du système ou épuisement des ressources de calcul, puis effectuer une analyse finale des données.

Il reste cependant une question importante à aborder : comment identifier les états conformationnels les moins échantillonnés? La réponse à cette question sera l'objet de la prochaine section.

L'analyse des simulations de dynamique moléculaire

Réduction de dimensionnalité des données

Une fois qu'un round d'échantillonnage des simulations est terminé, l'ensemble des états conformationnels les moins échantillonnés doit être déterminé. Or, même en se concentrant uniquement sur la protéine d'intérêt, dans le cas des GPCRs, le système contient facilement au-dessus de 250 résidus, chacun contenant entre 5 et 20 atomes, pour un total qui dépasse souvent les 4000 paramètres pour les positions des atomes et 1000 paramètres pour les angles dièdres en incluant ceux des chaînes latérales. De plus, chaque round de simulations apporte environ 32000 conformations différentes supplémentaires au cumul des simulations existantes. Que ce soit pour l'humain ou l'ordinateur, la quantité d'informations est trop importante pour qu'une analyse du système entier donne des résultats de bonne qualité dans un temps raisonnable. La solution à ce problème est de faire

appel à un ou plusieurs algorithmes de réduction de dimensionnalité, de façon à produire une représentation simplifiée du système qui contient l'information d'intérêt pour les analyses subséquentes.

Les algorithmes de réduction de dimensionnalité fonctionnent via l'optimisation d'une application des données brutes à un espace de dimension réduite (Sorzano *et al* , 2014) :

$$A := \mathbb{R}^n \rightarrow \mathbb{R}^m, m < n$$

$$\mathbf{r} \rightarrow A(\mathbf{r}) \stackrel{\text{def}}{=} \boldsymbol{\chi}$$

Les vecteurs $\boldsymbol{\chi}$ sont les vecteurs contenant les attributs d'intérêt de \mathbf{r} . Une des différentes formes que peut prendre cette application est celle d'une projection des données sur un ensemble de vecteurs de même dimension (vecteurs \mathbf{v} de module 1) :

$$A(\mathbf{r}) = V^T \mathbf{r}, V = (\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_m)$$

$$A(\mathbf{r}) = \begin{pmatrix} \mathbf{v}_1 \cdot \mathbf{r} \\ \mathbf{v}_2 \cdot \mathbf{r} \\ \vdots \\ \mathbf{v}_m \cdot \mathbf{r} \end{pmatrix}$$

Ces vecteurs sont choisis afin de mettre en emphase certains aspects des données étudiées. Par exemple, dans le cas de la PCA, les données de départ (\mathbf{x}) sont centrées sur leur moyenne ($\boldsymbol{\mu}$) et les vecteurs de la matrice de projection sont les vecteurs propres de la matrice des covariances des données (Σ) :

$$(7) \quad \boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$$

$$\Sigma = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$$

$$V \Lambda V^T = \Sigma$$

$$V^T = V^{-1}$$

Λ est une matrice contenant les valeurs propres associées aux vecteurs de V . Dans ce cas, il s'agit de la variance des données si on les projetait sur le vecteur correspondant (Sorzano *et al* , 2014). Ainsi, par exemple, si l'information d'intérêt était les changements conformationnels de grande amplitude d'une protéine, la PCA serait particulièrement appropriée. Il suffirait de projeter les données des trajectoires calculées, normalisées par rapport à la moyenne des éléments des trajectoires, sur les m vecteurs de V dont la variance est la plus grande (regroupés dans la matrice V_m) pour obtenir une représentation réduite des données contenant l'essentiel des informations qui nous intéressent :

$$\chi = V_m^T(\mathbf{r} - \boldsymbol{\mu})$$

Les vecteurs de V ont également comme propriété intéressante d'être une combinaison linéaire des différentes grandeurs incluses dans les vecteurs χ . Ainsi, les vecteurs propres de la matrice des covariances, ou de toute autre matrice obtenue de façon similaire, contiennent des informations sur les variations conjointes des grandeurs des vecteurs χ . D'autres méthodes de réduction de dimensionnalité de ce type existent, mais dans ce travail, celle qui nous intéresse le plus est la tICA. L'objectif de la tICA, contrairement à la PCA, n'est pas de se concentrer sur les mouvements de grande amplitude, mais plutôt sur ceux qui surviennent le moins fréquemment. Ainsi, le mouvement chaotique d'une chaîne latérale d'arginine exposée à l'eau ne sera pas pris en compte par la tICA, mais pourrait l'être par la PCA étant donné les grandes variations de position et d'angle qu'implique ce mouvement. Pour ce faire, la tICA, n'utilise pas la covariance, mais plutôt l'autocorrélation. Pour une trajectoire $\mathcal{T} = \{(\mathbf{r}_1, \mathbf{v}_1), \dots, (\mathbf{r}_N, \mathbf{v}_N)\}$, l'autocorrélation représente la similitude d'un élément de la trajectoire \mathbf{r}_i avec un autre $\mathbf{r}_{i+\tau}$ décalé dans le temps, les deux normalisés par rapport à la moyenne des éléments de la trajectoire. Pour comprendre l'utilité de l'autocorrélation dans l'étude des systèmes biologiques, il est utile d'introduire les processus markoviens. Un exemple de processus markovien continu est justement l'ensemble des solutions à l'équation de Langevin dont il a été question plus tôt (Gardinier, 2009), d'où l'utilité de les comprendre et de les étudier.

Processus markoviens et tICA

Un processus markovien est un processus dit « sans mémoire ». C'est-à-dire que pour une suite de variables aléatoires X_i dépendantes l'une de l'autre par leur évolution dans le temps, i.e. la variable aléatoire X_i qui évolue dans un certain processus pendant k étapes devient X_{i+k} , dans le cas d'un ensemble d'états accessibles S et de temps d'observation réguliers, un processus markovien satisfait à la propriété suivante (dite de processus sans mémoire) :

$$P(X_n = s_n | X_{n-1} = s_{n-1}, \dots, X_0 = s_0) = P(X_n = s_n | X_{n-1} = s_{n-1}), s_i \in S$$

Cette propriété est aussi valide, avec certaines modifications mineures, pour des ensembles continus (Nickelsen, 2014). L'utilité de cette propriété devient évidente lorsqu'on considère un saut dans le temps de plusieurs étapes. Considérons par exemple la probabilité suivante pour des intervalles de temps discrets :

$$P(X_n = s_n | X_{n-2} = s_{n-2})$$

Par la définition des probabilités conditionnelles, cette expression équivaut à :

$$P(X_n = s_n | X_{n-1} = s_{n-1}, X_{n-2} = s_{n-2})P(X_{n-1} = s_{n-1} | X_{n-2} = s_{n-2})P(X_{n-2} = s_{n-2})$$

Calculer une probabilité conditionnelle à deux variables est beaucoup moins trivial que calculer une probabilité conditionnelle à une seule, car le nombre de combinaisons d'états à considérer croît de façon exponentielle. Or, par la propriété de processus sans mémoire, cette expression se réduit à :

$$P(X_n = s_n | X_{n-1} = s_{n-1})P(X_{n-1} = s_{n-1} | X_{n-2} = s_{n-2})P(X_{n-2} = s_{n-2})$$

Cette propriété est d'autant plus puissante lorsqu'elle est combinée à l'équation de Chapman-Kolmogorov. Pour expliquer correctement cette équation dans le contexte qui

nous intéresse, passons au niveau d'abstraction supérieur d'un processus continu : une suite continue de variables aléatoires continues.

Par le même raisonnement, on définit la fonction de densité conditionnelle :

$$f_{X_{t+\tau}}(y|X_t = x)$$

$$\int_{-\infty}^a f_{X_{t+\tau}}(y|X_t = x) dy = P(X_{t+\tau} \leq a | X(t) = x)$$

L'équation de Chapman-Kolmogorov s'énonce alors ainsi :

$$\int_S f_{X_{t+\tau_1+\tau_2}}(z|X_{t+\tau_1} = y) f_{X_{t+\tau_1}}(y|X_t = x) dy = f_{X_{t+\tau_1+\tau_2}}(z|X_t = x)$$

Elle est également valide pour les processus multidimensionnels (décrits prochainement par des variables en caractères gras). Cette équation peut être interprétée de deux façons. D'abord, pour obtenir la probabilité conditionnelle de l'évolution d'un processus markovien, il suffit de marginaliser tous les états intermédiaires du processus (i.e. calculer la moyenne de la grandeur recherchée pour toutes les possibilités de ces états). Ensuite, et c'est là le plus important, cela signifie que l'évolution d'un système peut être décrite par une série de produits matriciels ou leur équivalent via l'opérateur de propagation (ou le propagateur). Pour un processus markovien continu, le propagateur relie la fonction $g(\mathbf{y})$ qui décrit le système observé au temps t à la fonction $h(\mathbf{y})$ qui décrit le même système au temps $t + \tau$ par l'expression suivante (Noé et Nüske, 2013) :

$$h(\mathbf{y}) = \int_{\mathbb{R}^n} f_{X_{t+\tau}}(\mathbf{y}|\mathbf{X}_t = \mathbf{x}) g(\mathbf{x}) d\mathbf{x} = \mathcal{P}(\tau) g(\mathbf{x})$$

L'opérateur $\mathcal{P}(\tau)$ est, dans le cas d'une variable aléatoire discrète (i.e. d'un nombre dénombrable d'états), une matrice de transition qui dicte la probabilité de passer d'un état à l'autre dans l'intervalle de temps τ , $g(\mathbf{x})$ devient un vecteur et l'application du propagateur devient un simple produit matriciel, mais les concepts associés restent les mêmes. Il suffit

de considérer $g(\mathbf{x})$ comme le vecteur, $f_{X_{t+\tau}}(\mathbf{y}|\mathbf{X}_t = \mathbf{x})$ comme la matrice et l'intégrale de leur produit comme le produit matriciel pour constater l'analogie. L'opérateur $\mathcal{P}(\tau)$, pour les processus Markoviens qui nous intéressent ici, admet une distribution stationnaire (à l'équilibre) $\mu(\mathbf{x})$ et un produit scalaire correspondant (Noé et Nüske, 2013; Nickelsen, 2014) :

$$\begin{aligned}\mathcal{P}(\tau)\mu(\mathbf{x}) &= \mu(\mathbf{x}) \\ \langle u|v \rangle &= \int_{\mathbf{r} \in \Omega} \frac{1}{\mu(\mathbf{x})} u(\mathbf{x})v(\mathbf{x})d\mathbf{x}\end{aligned}$$

Autrement dit, $\mu(\mathbf{x})$ est une fonction propre (ou *eigenfunction*) de $\mathcal{P}(\tau)$ avec comme valeur propre 1, ou encore, dans un contexte thermodynamique, la distribution des états du système à l'équilibre. Étant donné la relation entre l'intégrale et le produit matriciel dans ce contexte, il est naturel de créer un parallèle entre les vecteurs propres et valeurs propres d'une matrice. D'autres fonctions propres existent, toutes indépendantes du point de vue du produit scalaire et toutes ayant une valeur propre entre 1 et -1. Comme un processus markovien est sans mémoire, l'identité suivante est valide et découle directement de l'équation de Chapman-Kolmogorov (Noé et Nüske, 2013) :

$$\begin{aligned}& \int_{\mathbb{R}^n} f_{X_{t+\tau_1+\tau_2}}(\mathbf{z}|\mathbf{X}_t = \mathbf{x})g(\mathbf{x})d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f_{X_{t+\tau_1+\tau_2}}(\mathbf{z}|\mathbf{X}_{t+\tau_1} = \mathbf{y})f_{X_{t+\tau_1}}(\mathbf{y}|\mathbf{X}_t = \mathbf{x})g(\mathbf{x})d\mathbf{y}d\mathbf{x} \\ &= \mathcal{P}(\tau_2) \left(\mathcal{P}(\tau_1)g(\mathbf{x}(t)) \right) = \mathcal{P}(\tau_1 + \tau_2)g(\mathbf{x}(t))\end{aligned}$$

Et donc nécessairement, pour les valeurs propres de $\mathcal{P}(\tau)$ plus petites que 1 et supérieures à 0, les valeurs propres décroissent de façon exponentielle (Noé et Nüske, 2013) :

$$\begin{aligned}\lambda_i(\tau) &= \exp\left(-\frac{\tau}{t_i}\right) \\ -\frac{\tau}{\ln(\lambda_i(\tau))} &= t_i\end{aligned}$$

Les valeurs t_i sont appelées « échelles de temps » du système. Il s'ensuit que n'importe quelle approximation des fonctions propres du système a elle aussi une échelle de temps associée, et que cette échelle de temps est directement reliée à l'autocorrélation. Dans ce système, la densité de corrélation $C(\mathbf{x}, \mathbf{y}, \tau)$ et la fonction d'autocorrélation (acf) sont définis ainsi (Noé et Nüske, 2013):

$$\begin{aligned}
 C(\mathbf{x}, \mathbf{y}, \tau) &= \mu(\mathbf{x}) f_{X_{t+\tau}}(\mathbf{y} | \mathbf{X}_t = \mathbf{x}) \\
 acf(f, t, \tau) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{1}{\mu(\mathbf{x})} \mu(\mathbf{y}) f_{X_{t+\tau}}(\mathbf{y}) C(\mathbf{x}, \mathbf{y}, \tau) f_{X_t}(\mathbf{x}) d\mathbf{x} d\mathbf{y} \\
 &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{1}{\mu(\mathbf{x})} \mu(\mathbf{y}) f_{X_{t+\tau}}(\mathbf{y}) f_{X_{t+\tau}}(\mathbf{y} | \mathbf{X}_t = \mathbf{x}) \mu(\mathbf{x}) f_{X_t}(\mathbf{x}) d\mathbf{x} d\mathbf{y} \\
 &= \langle \mathcal{P}(\tau) \mu f | \mu f \rangle \stackrel{\text{def}}{=} \langle \mu f | \mathcal{P}(\tau) | \mu f \rangle
 \end{aligned}$$

Comme les fonctions propres l_i de $\mathcal{P}(\tau)$ sont orthogonales du point de vue du produit scalaire (i.e. $\langle l_p | l_q \rangle = \delta_{pq}$, le delta de Kronecker) et que n'importe quelle fonction f accessible par le système peut s'exprimer par une combinaison linéaire des fonctions propres l_i du système, la fonction normalisée par $\frac{1}{\mu}$ correspondante possède une échelle de temps associée (Noé et Nüske, 2013) :

$$\begin{aligned}
 f &= \sum_{k=1}^{\infty} \langle f | l_k \rangle l_k, l_1 = \mu, g = \frac{f}{\mu} \\
 \langle \mu g | \mathcal{P}(\tau) | \mu g \rangle &= \left\langle \sum_{k=1}^{\infty} \langle f | l_k \rangle \mathcal{P}(\tau) l_k \left| \sum_{k=1}^{\infty} \langle f | l_k \rangle l_k \right. \right\rangle = \left\langle \sum_{k=1}^{\infty} \langle f | l_k \rangle \lambda_k(\tau) l_k \left| \sum_{k=1}^{\infty} \langle f | l_k \rangle l_k \right. \right\rangle \\
 &= \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} \langle f | l_p \rangle \langle f | l_q \rangle \lambda_p(\tau) \langle l_p | l_q \rangle = \sum_{k=1}^{\infty} \langle f | l_k \rangle^2 \lambda_k(\tau) = \lambda_f(\tau) \\
 -\frac{\tau}{\ln(\lambda_f(\tau))} &= t_f, 1 > \lambda_f(\tau) > 0
 \end{aligned}$$

De même, il advient que chaque fonction propre normalisée par $\frac{1}{\mu}$ du système a une autocorrélation correspondant à sa valeur propre. L'autocorrélation est donc en quelque sorte une mesure de la qualité de l'approximation d'une distribution à l'équilibre du système par une fonction quelconque. En effet, une forte autocorrélation correspond à une échelle de temps élevée et une distribution des états qui varie plus lentement dans le temps. Concrètement, cela se traduit par deux cas possibles : les paramètres observés ne varient pas ou peu dans le temps, mais adoptent plusieurs valeurs distinctes (très probable) ou ces paramètres varient de façon périodique, avec une période régulière proche du délai du propagateur (très peu probable, de par la nature stochastique du système et du délai choisi). Or, d'un point de vue biologique, les changements d'états les plus intéressants observés en dynamique moléculaire sont aussi ceux qui sont les moins fréquents (Noé et Nüske, 2013; Pérez-Hernandez *et al*, 2013; Schwantes et Pande, 2013).

Pour revenir à la tICA, en sachant ce que représente l'autocorrélation, il est maintenant clair pourquoi cette technique de réduction de dimensionnalité est particulièrement appropriée pour l'analyse de systèmes biologiques comme les GPCRs. Dans le contexte de l'analyse de simulations de dynamique moléculaire, l'autocorrélation est définie comme une matrice et les fonctions propres et valeurs propres comme des solutions d'un problème de vecteurs propres et valeurs propres généralisé. La matrice d'autocorrélation de la trajectoire avec un délai τ vaut :

$$C(\tau) = \frac{1}{N - \tau} \sum_{\mathbf{r}(t_i) \in \mathcal{T}} (\mathbf{r}(t_i) - \boldsymbol{\mu})(\mathbf{r}(t_{i+\tau}) - \boldsymbol{\mu})^T$$

Les grandeurs des vecteurs \mathbf{r} peuvent être n'importe quoi, mais dans ce travail, il s'agit d'angles dièdres ajustés pour éliminer les artefacts de périodicité. Pour plusieurs trajectoires, cette somme est répétée sur leur ensemble, avec pour moyenne celle des éléments de toutes les trajectoires :

$$(8.1) \quad C(\tau) = \sum_{j=1}^n \left(\frac{1}{N_j - \tau} \sum_{\mathbf{r}(t_i) \in \mathcal{T}_j} (\mathbf{r}(t_i) - \boldsymbol{\mu})(\mathbf{r}(t_{i+\tau}) - \boldsymbol{\mu})^T \right)$$

La matrice $C_s(\tau)$ est ensuite calculée en effectuant la moyenne entre la matrice initiale calculée $C(\tau)$ et sa transposée, ce qui revient à considérer chaque trajectoire à l'endroit et à l'envers, et donne une matrice symétrique. Cette opération peut engendrer un biais dépendamment de la différence entre l'ensemble des trajectoires et la représentation du système à l'équilibre, mais il n'existe pas de solution évidente à ce problème (Schwantes et Pande, 2013). La réduction de dimensionnalité passe par l'identification des vecteurs propres généralisés qui satisfont aux équations suivantes (Schwantes et Pande, 2013), avec les mêmes définitions de Σ et Λ que pour la PCA, à l'exception que Λ contient maintenant des valeurs d'autocorrélation et non de variance comme dans les équations 6 :

$$(8.2) \quad C_s(\tau)V = \Sigma V \Lambda$$

$$V^T \Sigma V = I$$

Comme Σ est une matrice définie positive et que les deux matrices sont symétriques, il existe une solution exacte et d'expression (relativement) simple au problème :

$$C'_s(\tau) = \Lambda_{\Sigma}^{-\frac{1}{2}} V_{\Sigma}^T C_s(\tau) V_{\Sigma} \Lambda_{\Sigma}^{-\frac{1}{2}}$$

$$V = V_{\Sigma} \Lambda_{\Sigma}^{-\frac{1}{2}} V_{C'_s(\tau)} \quad \Lambda = \Lambda_{C'_s(\tau)}$$

Les matrices V_{Σ} et Λ_{Σ} sont celles de la décomposition propre de Σ ($V_{\Sigma} \Lambda_{\Sigma} V_{\Sigma}^T = \Sigma$), de même pour $V_{C'_s(\tau)}$, $\Lambda_{C'_s(\tau)}$ et $C'_s(\tau)$. Les modules des vecteurs propres obtenus permettent de compenser la contribution de la variance et de la covariance aux valeurs d'autocorrélation, de façon à donner des valeurs propres entre -1 et 1. La réduction de dimensionnalité s'effectue alors en choisissant les m vecteurs présentant la plus grande autocorrélation :

$$\chi = V_m^T (\mathbf{r} - \boldsymbol{\mu})$$

Ainsi, les changements conformationnels peu fréquents devraient être plus visibles si les données brutes sont projetées sur les vecteurs de la tICA que ceux de la PCA ou d'autres techniques similaires.

Maintenant que chaque élément de chaque trajectoire est de dimensionnalité réduite, en général moins de 10 dimensions par rapport aux 1000 à 4000 de départ, un seul problème persiste : la quantité de conformations différentes reste de l'ordre de la dizaine à la centaine de milliers. Pour réduire la taille du système et se concentrer sur ses conformations représentatives, une technique de groupage des données a dû être utilisée. Sa validité dans un contexte thermodynamique a pu être évaluée en construisant un modèle de Markov à partir des différents états identifiés.

Chaînes de Markov, k-means et mélange de gaussiennes

Le groupage des données consiste aussi en une réduction de dimensionnalité des données. Pour une collection de groupes de données (états) $S_i \in \mathbb{S}$, l'application prend cette forme :

$$A(\mathbf{r}) = (s_0, \dots, s_n), s_i = P(\mathbf{r} \in S_i)$$

L'avantage de cette technique est que l'ensemble des données du système peut maintenant être associé à un ensemble fortement réduit d'états dans lesquels les points qui en font partie sont similaire entre eux. Ainsi, au lieu de concentrer ses efforts sur l'analyse de dizaine de milliers de différentes conformations, il est maintenant possible de se concentrer sur au plus quelques centaines, voire quelques dizaines ou encore moins si certaines opérations subséquentes sont effectuées. Plusieurs types de groupages de données existent, mais ceux qui nous intéressent ici sont le groupage de type *k-means* et le groupage par mélange de gaussiennes. La raison est que ces deux types de groupages sont certes d'usage répandu, mais sont surtout utilisés en biochimie structurale pour la construction d'un MSM dans le cas du *k-means* et un HMM dans le cas du mélange de Gaussiennes (Harrigan *et al*, 2017), deux modèles utilisant la théorie des chaînes de Markov.

Une chaîne de Markov est une forme simplifiée d'un processus markovien, où le nombre d'états possible est dénombrable. Ainsi, l'opérateur $\mathcal{P}(\tau)$ devient une matrice et la fonction de densité de probabilité devient un vecteur contenant la probabilité de se situer dans chaque état au temps t :

$$\begin{aligned}\mathcal{P}(\tau) &\in \mathbb{R}^{n \times n} \\ (9) \quad \mathbf{s}(t + \tau) &= \mathcal{P}(\tau)\mathbf{s}(t) = \mathcal{P}(\tau) \begin{pmatrix} s_1(t) \\ \vdots \\ s_n(t) \end{pmatrix} \\ s_i(t) &= P(Y_t = S_i), i = 1 \dots n \\ \mathcal{P}(\tau) &= [P(Y_{t+\tau} = S_j | Y_t = S_i)]_{n \times n}\end{aligned}$$

Le produit scalaire et l'autocorrélation définis plus tôt deviennent leurs formes vectorielles simplifiées et une distribution stationnaire existe aussi. Il s'agit d'un vecteur propre (à droite, car $\mathcal{P}(\tau)$ n'est pas symétrique) de la matrice $\mathcal{P}(\tau)$:

$$\begin{aligned}\langle \mathbf{u} | \mathbf{v} \rangle &= \mathbf{u} \cdot \mathbf{v} \\ acf(\mathbf{s}) &= \mathbf{s}(t + \tau) \cdot \mathbf{s}(t) = \mathbf{s}^T(t) \mathcal{P}(\tau) \mathbf{s}(t) \\ \mathcal{P}(\tau) \mathbf{s}_{eq}(t) &= \mathbf{s}_{eq}(t) = \boldsymbol{\mu}\end{aligned}$$

La différence fondamentale entre chaîne de Markov et HMM repose dans la façon d'assigner les probabilités qu'un processus se trouve dans un état particulier. Dans le cas d'un MSM, l'état qu'adopte un processus à un temps déterminé est connu de façon absolue. C'est pourquoi dans ce cas l'algorithme de type *k-means* est utilisé :

$$P(Y_t = S_i | \mathbf{r}(t)) = \begin{cases} 1, & \mathbf{r}(t) \in S_i \\ 0, & \mathbf{r}(t) \notin S_i \end{cases}, S_i \subseteq \mathbb{R}^p$$

$\mathbf{r}(t) \in \mathbb{R}^p$ est ici un ensemble observations du processus défini par l'équation de Langevin. C'est précisément pour cette raison que la réduction de dimensionnalité est nécessaire à l'analyse par modèle de Markov : un $\mathbf{r}(t)$ ayant trop de dimensions devient difficile à assigner à un état.

Dans un contexte d'analyse de dynamique moléculaire, cela revient à partitionner, ou « paver », l'ensemble du système à l'aide d'un diagramme de Voronoi, où la distance entre un point de données et le centre d'un état $\mathbf{c}(S_i)$ dicte l'appartenance de ce point :

$$P(S_i, t | \mathbf{r}(t)) = \begin{cases} 1, & i = \min_j |\mathbf{r}(t) - \mathbf{c}(S_j)| \\ 0, & i \neq \min_j |\mathbf{r}(t) - \mathbf{c}(S_j)| \end{cases}$$

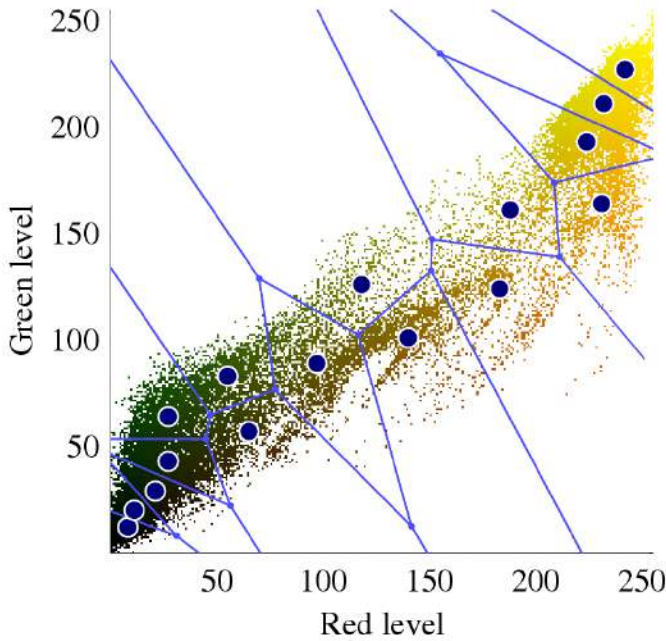


Figure 9 : Exemple de groupage par *k-means* sur les pixels d'une image

Chaque point bleu représente le centre d'un groupe et chaque ligne bleue sépare un groupe d'un autre. Le motif résultant est un pavage de Voronoi dont chaque tuile est un état.

Source de l'image :

https://commons.wikimedia.org/wiki/File:Rosa_Gold_Glow_2_small_noblue_color_space.png (image du domaine public)

Dans le cas d'un HMM, les états sont en réalité des fonctions de densité de probabilité multipliées par un facteur de « population » et la probabilité que le processus se retrouve dans un certain état est interprétée comme une fonction des différentes fonctions de densité de chaque état avec la formule de Bayes et le développement de la probabilité marginale $P(\mathbf{r}(t))$ (Bishop, 2006) :

$$s_i = P(S_i | \mathbf{r}, t)$$

$$P(Y_t = S_i | \mathbf{r}(t)) = \frac{P(\mathbf{r}(t) | Y_t = S_i) P(Y_t = S_i)}{P(\mathbf{r}(t))} = \frac{P(\mathbf{r}(t) | Y_t = S_i) P(Y_t = S_i)}{\sum_{i=1}^n P(\mathbf{r}(t) | Y_t = S_i) P(Y_t = S_i)}$$

$$\propto \frac{\mu_{S_i} f_X(\mathbf{r}(t) | \boldsymbol{\mu}_i, \Sigma_i)}{\sum_{i=1}^n \mu_{S_i} f_X(\mathbf{r}(t) | \boldsymbol{\mu}_i, \Sigma_i)}, \sum_{i=1}^n \mu_{S_i} = 1$$

En effet, la grandeur $P(\mathbf{r}(t) | Y_t = S_i)$ est proportionnelle à la probabilité qu'une trajectoire quelconque $\mathbf{s}(t')$ passe par une petite région autour de $\mathbf{r}(t)$ à un temps quelconque. Dans le cas d'un mélange de Gaussiennes, l'expression de $P(\mathbf{r}(t) | Y_t = S_i)$ devient donc $f_X(\mathbf{r}(t) | \boldsymbol{\mu}_i, \Sigma_i)$, qui représente ici une loi normale multivariée de moyenne $\boldsymbol{\mu}_i$ et de matrice des covariances Σ_i . C'est-à-dire :

$$(10) \quad f_X(\mathbf{r}(t) | \boldsymbol{\mu}_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_i|}} \exp\left(-\frac{1}{2}(\mathbf{r}(t) - \boldsymbol{\mu}_i)^T \Sigma_i^{-1} (\mathbf{r}(t) - \boldsymbol{\mu}_i)\right)$$

$\boldsymbol{\mu}_i$ représente ici la portion de la distribution stationnaire du système correspondant à l'état S_i et $|\Sigma_i|$ représente le déterminant de Σ_i . Par souci de simplicité et de rapidité de calcul, le mélange de Gaussiennes est souvent restreint à l'utilisation de matrices des covariances diagonales, c'est-à-dire ne contenant aucun élément non-nul hors de la diagonale de la matrice (McGibbon *et al*, 2014).

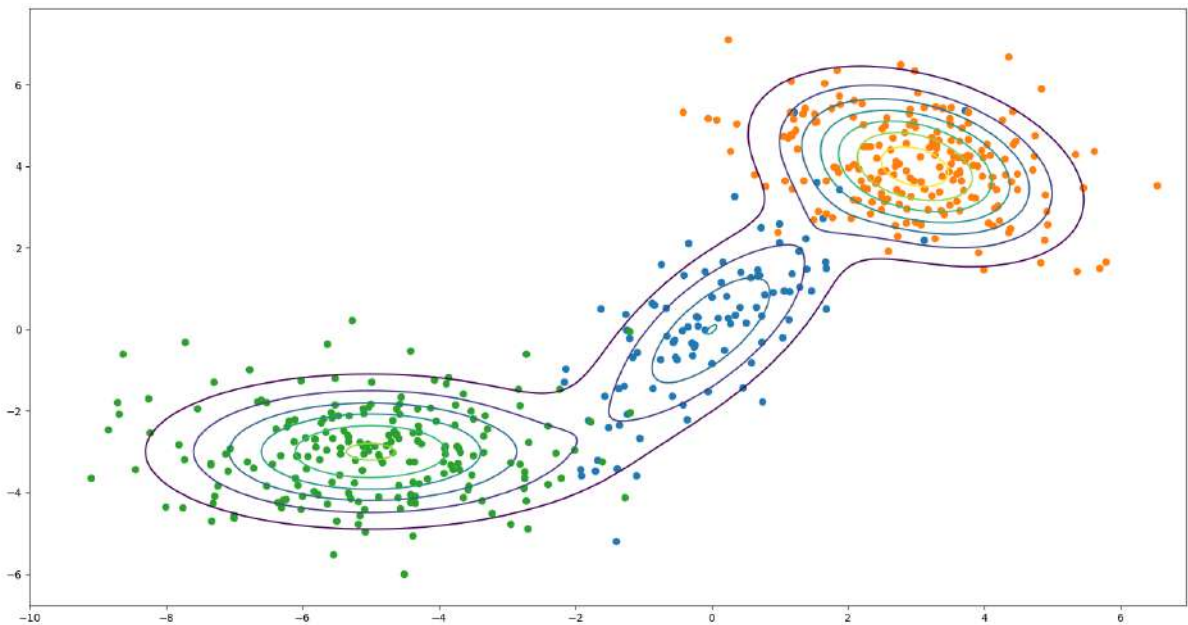


Figure 10 : Exemple de groupage par mélange de Gaussiennes

Ici, les points tirés de trois distributions Gaussiennes multivariées sont regroupés en trois états Gaussiens différents. La fonction de densité de probabilité résultante pour l'ensemble des points est une combinaison de ces trois Gaussiennes et est visible ici sous forme de courbes de niveau.

La différence entre MSM et HMM est illustrée de façon élégante par l'exemple de la « climatologie des crèmes glacées » (Eisner, 2002). Supposons que l'on est en l'an 4018 et qu'on cherche à construire une chaîne de Markov à partir des données météo de Sherbrooke en 2018 en regroupant l'ensemble des journées dans trois états : jour ensoleillé, jour nuageux et jour pluvieux. Si on dispose des données météorologiques brutes, il est trivial d'établir des critères pour ce qu'est un jour ensoleillé, nuageux ou pluvieux, de répartir les jours dans chaque état et de calculer les probabilités de transition d'un MSM. Cependant, imaginons que, pour une quelconque raison, les données météorologiques brutes sont perdues, mais que le journal d'un résident de Sherbrooke de la même année existe encore et a été retrouvé. Imaginons maintenant que ce journal contient pour chaque journée de l'été le nombre de crèmes glacées qu'a mangées ce résident. Pour construire une chaîne de Markov, l'opération est beaucoup moins triviale : il faut associer une distribution des probabilités de manger une certaine quantité de crèmes glacées en fonction du type de journée, puis optimiser cette distribution et les probabilités de transition de la chaîne de Markov de façon à trouver à la fois la chaîne de Markov et la succession d'états la plus probable en fonction des données brutes. La différence entre les deux est illustrée dans la figure 11.

Dans le contexte de notre analyse, un GMHMM a été choisi pour plusieurs raisons. D'abord, le groupage *k-means*, de par sa nature, constitue une approximation de la distribution à l'équilibre du système par un ensemble de fonctions indicatrices de différents poids, donc une approximation « en escalier », donc différentiable en presque tout point et de dérivée nulle partout, d'une fonction différentiable en tout point et de dérivée non-nulle presque partout (Noé et Nüske, 2013). Ce type d'approximation requiert une quantité importante de fonctions indicatrices, soit des milliers, voire des dizaines de milliers, dépendant de la dimensionnalité du système (McGibbon *et al* , 2014). En effet, l'approximation par fonctions indicatrices est excessivement sensible au phénomène connu sous le nom de « malédiction de la dimensionnalité » (ou *curse of dimensionality* en

anglais). C'est-à-dire que, considérant que 10 fonctions indicatrices sont suffisantes pour approximer une fonction, disons une gaussienne, en une dimension, il faudra plutôt 100 fonctions indicatrices pour approximer la fonction étendue en deux dimensions avec la même précision, 1000 en trois, etc. Dans le cas d'une approximation par gaussiennes, comme dans un HMM, la quantité d'états augmente beaucoup moins rapidement, voire pas du tout, puisqu'il s'agit d'ajouter une dimension à la fonction de densité pour pallier au problème, ce qui facilite grandement l'interprétation du système au moment d'en faire l'analyse.

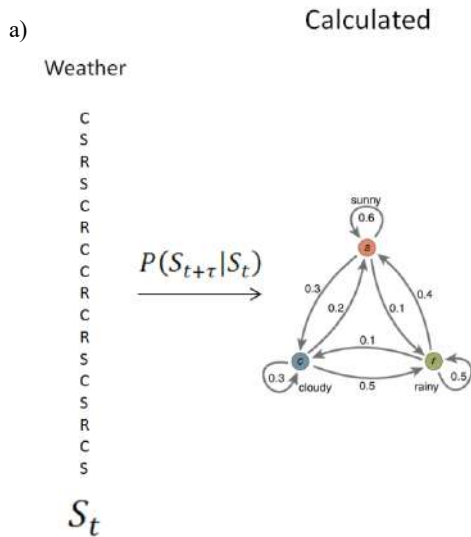
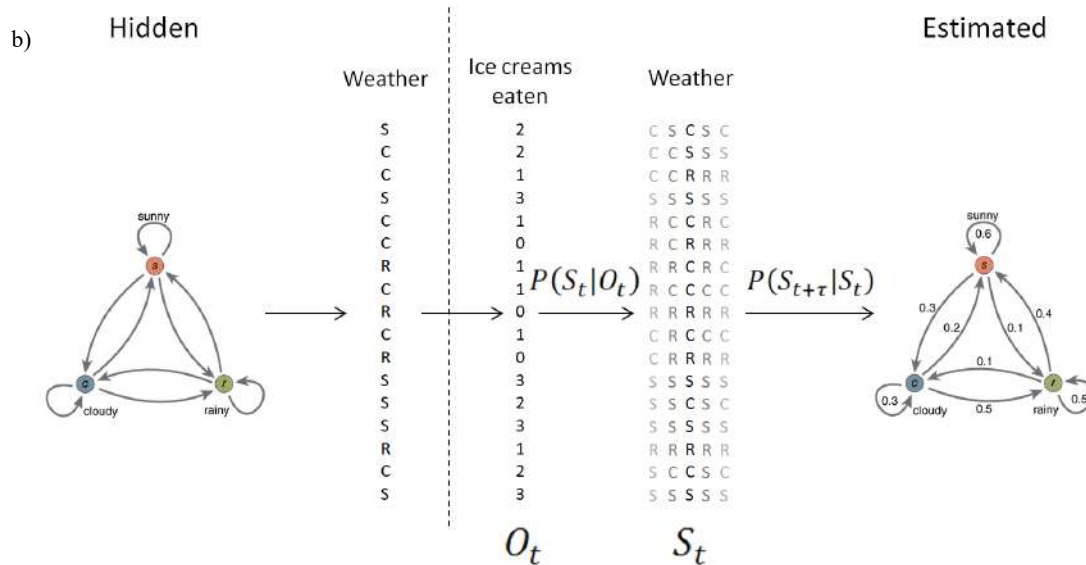


Figure 11 : Comparaison d'une chaîne de Markov et d'un HMM

a) Une chaîne de Markov peut être construite lorsqu'on connaît exactement les états correspondant aux observations de l'évolution d'un système, ainsi que la succession de ces états. Il suffit de calculer les probabilités de succession $P(S_{t+\tau}|S_t)$ pour obtenir la chaîne correspondante. b) Un HMM peut être construit lorsque les observations ne peuvent être attribuées avec certitude à un état ou un autre. Les probabilités d'association des observations aux états $P(S_t|O_t)$ doivent être optimisées simultanément aux probabilités de succession afin d'obtenir une estimation adéquate de la chaîne de Markov.



Ensuite, la construction d'un MSM présente certains désavantages en ce qui a trait à l'optimisation de modèle. En effet, dans ce cas, l'utilisation d'un algorithme de type *k-means* consiste en un pavage de l'espace qui contient les données et fait ainsi partie des étapes de pré-calcul de la chaîne de Markov, puisque l'assignation des différents points de données à un état est préalable et nécessaire à sa construction (McGibbon *et al* , 2014). Dans le cas d'un HMM, l'optimisation des différents paramètres utilisés pour chaque état n'est pas nécessaire, ni préalable à la construction du modèle, puisque l'assignation à un état n'est pas absolue. L'appartenance d'un point de données à un état peut donc être optimisée en même temps que la chaîne de Markov elle-même en utilisant une variante de la plausibilité. La plausibilité d'un HMM pour une série de données $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, une série d'états $S = \{S_1, \dots, S_n\}$, la plus probable étant donné la série X et les paramètres

des fonctions de densité de probabilité correspondantes, représentés par $\Theta = \{\theta_1, \dots, \theta_n\}$, vaut (Bishop, 2006; McGibbon *et al* , 2014) :

$$\begin{aligned} \mathcal{L}(X|S, \Theta) &= P(S_1) \prod_{i=1}^{n-1} P(Y_{i+1} = S_{i+1} | Y_i = S_i) \prod_{i=1}^n P(Y_i = S_i | \theta_i) \\ &= \mu_{S_1} \prod_{i=1}^{n-1} P(Y_{i+1} = S_{i+1} | Y_i = S_i) \prod_{i=1}^n f_X(\mathbf{r}(t) | \boldsymbol{\mu}_i, \Sigma_i) \end{aligned}$$

Cette expression contient à la fois le propagateur de la chaîne de Markov, sous la forme des $P(Y_{i+1} = S_{i+1} | Y_i = S_i)$ et de la distribution stationnaire $\boldsymbol{\mu}$, et les paramètres des fonctions de densité de probabilité, sous la forme de $P(Y_i = S_i | \theta_i)$ ou $f_X(\mathbf{r}(t) | \boldsymbol{\mu}_i, \Sigma_i)$, ce qui permet de maximiser la plausibilité du système en optimisant simultanément les paramètres des distributions normales et celles de la matrice de transition, tout cela, bien sûr, moyennant quelques ajustements pour assurer le réalisme thermodynamique du modèle, une bonne répartition des états, et la présence d'une distribution à l'équilibre du système (McGibbon *et al* , 2014).

Finalement, un HMM a été choisi pour la facilité avec laquelle il est possible de regrouper ses différents états par similarité structurale et cinétique. En effet, comme un mélange de gaussiennes requiert beaucoup moins d'états qu'un pavage via *k-means*, les vecteurs propres peuvent être utilisés directement pour identifier les différents états d'un système, sans avoir à utiliser des techniques subséquentes comme la PCCA ou ses variantes (Röblitz et Weber, 2013). De plus, pour la même raison, la classification selon des similitudes conformationnelles est aussi assez simple. Comme la densité spatiale des états dans un mélange de gaussiennes est très faible par rapport à celle d'un pavage par *k-means*, il est possible de regrouper les états selon leurs similitudes structurales en utilisant des mesures de distance statistique, comme la distance de Bhattacharrya (Bhattacharrya, 1943; Choi et Lee, 2003), sans obtenir un ou deux grands groupes qu'il faudrait ensuite séparer en utilisant d'autres techniques. Cette raison du choix d'un HMM est expliquée plus en détail dans la prochaine section.

Considérations pratiques et méta-analyse

Jusqu'à maintenant, le protocole choisi d'échantillonnage et d'analyse du système par dynamique moléculaire comprend les composantes suivantes :

1. Effectuer 8 simulations en parallèle du système avec la même conformation de départ, mais des vitesses différentes
2. Identifier les états conformationnels les moins échantillonnés
 - a. Projeter les données conformationnelles brutes sur les vecteurs propres de haute autocorrélation obtenus par tICA.
 - b. Grouper les données et construire une chaîne de Markov en optimisant un GMHMM.
3. Effectuer 8 autres simulations avec les états identifiés en 2 et en réinitialisant les vitesses des atomes de façon aléatoire
4. Répéter 2 et 3 jusqu'à convergence du système ou épuisement des ressources de calcul, puis effectuer une analyse finale des données.

Malgré les progrès effectués, trois questions demeurent :

- Quelles données de simulation utiliser : les positions ou les angles dièdres?
- Comment optimiser la dimensionnalité des données et le nombre d'états de façon à maximiser la représentativité thermodynamique du système?
- Est-il possible de tirer davantage d'information sur la nature des changements conformationnels observés à partir des informations tirées directement de la tICA et du HMM?

En effet, dans les sections précédentes, il a été mentionné de réduction de dimensionnalité sous forme de projection sur vecteurs et groupage, mais dans le contexte de l'analyse de simulations de dynamique moléculaire, ce qui est aussi important est de savoir quel type de données utiliser. Il a aussi été jusqu'à maintenant question des raisons pour lesquelles la tICA et le HMM sont utilisés à la place d'autres alternatives, mais il n'a jamais été précisé combien de vecteurs propres de la tICA ou de gaussiennes utiliser pour obtenir un système aussi représentatif que possible du véritable paysage énergétique du processus. De plus, bien qu'un vecteur propre de la tICA consiste en une représentation d'un changement conformationnel lent, il n'y a aucune information sur l'importance d'un résidu ou angle dièdre en particulier pour l'occurrence dudit mouvement. Le vecteur propre ne donne que l'amplitude de la variation de chaque paramètre. Pour le HMM, le problème est similaire : les données ont effectivement été classifiées dans différents états, mais l'existence d'états différents ne signifie en rien que les données contenues dans ces états sont totalement différentes, ni quels changements conformationnels les séparent. Ces différentes questions et problématiques sont l'objet de cette section.

D'abord, pour le choix des données, le choix des angles dièdres présente certains avantages, les deux plus importants étant une représentation réduite du système et une représentation indépendante de la position et de l'orientation de la protéine. En effet, la représentation d'un résidu par ses positions demande $3n$ dimensions, n étant ici le nombre d'atomes prenant part à un lien ayant au moins un degré de liberté rotationnel indépendant des autres atomes sélectionnés (i.e. seuls deux atomes d'un groupement aromatique sont requis). Pour les angles dièdres, la représentation demande au plus six dimensions pour une

lysine et un minimum de deux pour la glycine : les angles du squelette de la protéine, φ et ψ pour la glycine et la lysine, et les angles de la chaîne latérale, χ_1 , χ_2 , χ_3 et χ_4 pour la lysine. De plus, la représentation de la protéine par angles dièdres est indépendante des translations et rotations de la protéine lors des simulations et ne requiert donc pas d'alignement préalable des structures avant l'analyse. Un désavantage important de la représentation de la protéine par des angles dièdres est la nécessité d'une transformation des données qui apporte une distorsion aux variations observées. En effet, la simple représentation de la protéine par des angles n'est pas suffisante : le passage d'un angle dièdre de 179° à -179° ne constitue pas une variation de 358° , mais seulement 2° . Pour remédier à ce problème, le sinus ou le cosinus des angles dièdres sont utilisés pour représenter la protéine en raison de leur périodicité. Comme la grande majorité des angles φ de la protéine se situent autour de 90° , le cosinus est la fonction choisie pour cet angle. De même, comme la majorité des angles ψ de la protéine se trouvent autour de 0° et 180° et que les angles χ se trouvent majoritairement autour de 60° , -60° et 180° , le sinus est la fonction choisie (Laskowski *et al*, 1993). Cela permet de maximiser le déterminant du Jacobien autour des minima énergétiques des angles dièdres. En effet, la transformation des angles du système en leur sinus ou cosinus apporte une distorsion aux données : le développement de Taylor du premier degré en un point de cet espace est exprimé en fonction du Jacobien. Le déterminant du Jacobien permet une estimation de la contraction de l'espace autour du point :

$$f(x) \approx f(c) + J(f, c)(x - c)$$

$$J(f, c) = \left[\frac{\partial}{\partial x_j} f_i(x), x = c \right]_{6 \times 6} = \begin{bmatrix} -\sin(\varphi) & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos(\psi) & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(\chi_1) & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cos(\chi_4) \end{bmatrix}_{6 \times 6}$$

$$\det J(f, c) = -\sin(\varphi)\cos(\psi) \prod_{i=1}^4 \cos(\chi_i)$$

Le déterminant présente des maxima (en valeur absolue) à $\varphi = \pm 90^\circ$ et aux autres angles à 0° ou 180° , ce qui indique que la distorsion des données est minimale dans ces régions, et

la fonction de départ change de signe pour les angles dièdres à -60° par rapport à ceux à 60° , ce qui devrait permettre d'observer des changements conformationnels entre ces deux minima énergétiques. Ceci dit, malgré ce choix éclairé des fonctions utilisées pour contraindre l'espace de départ à la périodicité, des inconvénients subsistent. Par exemple, les changements conformationnels de $(\pm 90 - x)^\circ$ à $(\pm 90 + x)^\circ$ ne sont pas observables pour les angles ψ et χ , de même pour x° à $-x^\circ$ ou $(180 + x)^\circ$ à $(180 - x)^\circ$ pour l'angle φ . De plus, le Jacobien est nul autour de 0° et 180° pour φ et $\pm 90^\circ$ pour ψ et χ , ce qui empêche toute représentation fiable des données autour de ces régions. Ceci dit, comme ces régions sont de population minimale pour ces différents angles et qu'il n'y a pas de maxima de population répartis symétriquement autour de ces angles (Laskowski *et al*, 1993), ces inconvénients devraient avoir un impact minimal sur la qualité des données.

Une fois les données prêtes pour l'analyse, le problème est maintenant de savoir jusqu'où pousser l'analyse pour obtenir la représentation à la fois la plus compacte et la plus représentative possible du système. Pour ce faire, deux critères sont utilisés. Le premier critère sert à trouver la meilleure approximation disponible de la distribution à l'équilibre du système, étant donné une certaine dimensionnalité, et le second sert à optimiser la dimensionnalité des données de façon à inclure, autant que possible, la totalité des changements conformationnels qui permettent de décrire le système à l'équilibre de la façon la plus complète et la plus concise possible.

Le premier critère utilisé est le BIC (Wit *et al*, 2012). Ce critère est utilisé pour optimiser le HMM du système. Pour un modèle issu de n points de données et contenant K états représentés par p paramètres, le BIC s'exprime ainsi :

$$BIC(S, \Theta|X) = -2\mathcal{L}(X|S, \Theta) + pK \log(n)$$

Dans notre cas, le BIC vaut, pour un modèle de dimension N :

$$BIC(S, \Theta|X) = -2 \log(\mathcal{L}(X|S, \Theta)) + 2NK \log(n)$$

$\mathcal{L}(X|S, \Theta)$ est tel que décrit plus haut dans l'optimisation des HMMs. La minimisation du BIC permet d'identifier le modèle le plus probable considérant les données disponibles (Wit *et al*, 2012), et ainsi d'identifier la meilleure approximation de la distribution à l'équilibre du système. Pour identifier le meilleur modèle, le nombre d'état est augmenté graduellement jusqu'à l'atteinte d'un plateau. C'est au début de ce plateau que le modèle le plus concis parmi les meilleurs se trouve. La connaissance de la distribution à l'équilibre issue de ce modèle est essentielle pour l'utilisation du second critère.

Le second critère utilisé pour optimiser le modèle se concentre sur la dimensionnalité des données. Ce critère se base sur les échelles de temps de la chaîne de Markov issue des données. Il s'agit du principe variationnel des processus Markoviens. Considérant un système dont la distribution stationnaire μ est connue et une fonction orthogonale à la distribution à l'équilibre du système en vertu du produit scalaire, si cette fonction n'est pas égale à la seconde fonction propre du propagateur, alors son autocorrélation est plus petite que celle de la seconde fonction propre du système (Noé et Nüske, 2013) :

$$\begin{aligned}
 f &= \sum_{k=1}^{\infty} \langle f | l_k \rangle l_k, l_1 = \mu, g = \frac{f}{\mu}, \langle f | \mu \rangle = 0, g \neq \frac{l_2}{\mu} \\
 \langle \mu g | \mathcal{P}(\tau) | \mu g \rangle &= \sum_{k=1}^{\infty} \langle f | l_k \rangle^2 \lambda_k(\tau) = \sum_{k=2}^{\infty} \langle f | l_k \rangle^2 \lambda_k(\tau) \\
 &< \sum_{k=2}^{\infty} \langle f | l_k \rangle^2 \lambda_2(\tau) = \lambda_2(\tau) \\
 \lambda_f(\tau) &< \lambda_2(\tau)
 \end{aligned}$$

Ainsi, pour optimiser la qualité du modèle du système, l'objectif est de maximiser les échelles de temps dudit modèle. Comme pour des systèmes de même dimension, le BIC est déjà utilisé, il serait légitime de se demander pourquoi ne pas l'utiliser pour comparer des systèmes de dimensions différentes entre eux. La raison pour laquelle il est impossible d'effectuer une telle comparaison est que les termes de la plausibilité associés aux lois normales diminuent systématiquement avec le nombre de dimensions. Par exemple, considérant une seule matrice gaussienne de moyenne $\mathbf{0}$ et matrice des covariances I , la

plausibilité d'un point à **0** pour un système de dimension $n + 1$ par rapport à celle d'un système de dimension n vaut (éq. 10) :

$$\frac{e^0}{(2\pi)^{\frac{(n+1)}{2}}} \cdot \frac{(2\pi)^{\frac{n}{2}}}{e^0} = \frac{1}{\sqrt{2\pi}}$$

Considérant que le système contient aussi N points de données, le facteur devient $(\sqrt{2\pi})^{-N/2}$ pour la plausibilité. De plus, considérant que les variances sont différentes, alors la variation de la plausibilité devient de plus en plus incertaine d'une dimensionnalité à l'autre. La façon la plus fiable de comparer deux systèmes de dimensionnalité différente est donc de comparer les échelles de temps. Ceci dit, comme la deuxième fonction propre identifiée par le modèle du système est une approximation de la véritable fonction propre, le principe variationnel ne peut être appliqué à la troisième fonction propre et au-delà (Pérez-Hernandez *et al*, 2013). L'échelle de temps la plus fiable hormis celle de la distribution stationnaire est donc la deuxième, et c'est celle qui est utilisée pour l'évaluation du modèle.

Le protocole d'échantillonnage final est donc le suivant :

1. Effectuer 8 simulations en parallèle du système avec la même conformation de départ, mais des vitesses différentes
2. Identifier les états conformationnels les moins échantillonnés
 - a. Transformer les données brutes d'angles dièdres en leur sinus ou cosinus
 - b. Projeter les données conformationnelles brutes sur le premier vecteur propre de haute autocorrélation obtenus par tICA.
 - c. Grouper les données et construire une chaîne de Markov en optimisant un HMM avec mélange de Gaussiennes.
 - d. Répéter c en ajoutant un état au système jusqu'à optimisation du BIC.
 - e. Répéter b, c et d en ajoutant un vecteur propre de la tICA à chaque itération, jusqu'à optimisation de la seconde échelle de temps de la chaîne de Markov issue du système.

3. Effectuer 8 autres simulations avec les états identifiés en 2 et en réinitialisant les vitesses des atomes de façon aléatoire.
4. Répéter 2 et 3 jusqu'à convergence du système ou épuisement des ressources de calcul, puis effectuer une analyse finale des données en suivant le protocole de l'étape 2.

Maintenant, une fois l'échantillonnage terminé la question qui demeure est de savoir si les techniques d'analyse utilisées contiennent plus d'informations qu'elles n'en donnent d'emblée. Les deux informations dont le manque est le plus criant dans cette analyse sont les suivants :

- Quels résidus sont les plus importants pour les changements conformationnels représentés par les vecteurs propres de la tICA?
- Quels changements conformationnels permettent de séparer les états du HMM?

Une solution est proposée pour trouver chaque information à partir du modèle. Chaque solution est détaillée dans l'article associé à ce travail.

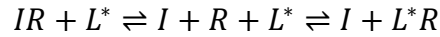
Présentation des méthodes expérimentales utilisées

Les essais de liaison par compétition

Les essais de liaison de compétition suivent un principe similaire aux essais par saturation (éq. 1), mais la mesure de l'affinité s'effectue de façon relative. Le ligand chaud est exposé à un ligand compétiteur, non-radiomarké (froid), qui lie le même site que le ligand radiomarké, ce qui le déloge et, par conséquent, diminue la concentration de ligand chaud lié aux récepteurs, donc aussi le compte des désintégrations observées. Le protocole de l'essai est pratiquement identique, mais il y a quelques différences majeures:

- La concentration du ligand chaud est constante partout
- La préparation de membranes et le ligand radiomarké sont exposés à des quantités croissantes du ligand froid.

L'équilibre de l'essai de liaison par saturation est donc modifié :



$$K_I = \frac{[I][R]}{[IR]}$$

$$K_d = \frac{[L^*][R]}{[L^*R]}$$

Les hypothèses de départ sont similaires à celles de l'essai par saturation :

$$\begin{aligned} [L^*]_{tot} &= [L^*] + [L^*R] \approx [L^*] \\ [R]_{tot} &= [R] + [IR] + [L^*R] \approx cste \\ [I]_{tot} &= [I] + [IR] \approx [I] \end{aligned}$$

De même que pour l'essai par saturation, en isolant $[RI]$ dans l'équation de K_I , en substituant ensuite l'expression de $[RI]$ dans l'équation de K_d , puis en isolant $[RL^*]$, on obtient :

$$\begin{aligned} (11) \quad \frac{\partial}{\partial t} N(t, [L^*], [I]) &= \frac{k_{rad}[R]_{tot}[L^*]}{K_d \left(1 + \frac{[I]}{K_I}\right) + [L^*]} = \frac{k_{rad}[R]_{tot}[L^*]}{K_d + [L^*]} \frac{1}{1 + \frac{[I]}{IC_{50}}} \\ IC_{50} &= K_I \left(1 + \frac{[L^*]}{K_d}\right) \end{aligned}$$

Cette relation entre IC_{50} , plus facile à mesurer, et le K_I de l'inhibiteur est dérivée de l'équation de Cheng-Prusoff (Cheng et Prusoff, 1973). À toutes fins pratiques, l'équation de gauche n'est pratiquement jamais utilisée. C'est plutôt l'équation de droite qui l'est, sous une forme semi-logarithmique simplifiée. Pour un bruit de fond N_{min} et une radioactivité maximale détectée N_{max} :

$$\frac{\partial}{\partial t} N(t, x) = N_{min} + \frac{N_{max} - N_{min}}{1 + 10^{x - \log IC_{50}}}, x = \log[I]$$

Ce type de courbe présente aussi un point d'inflexion à $x = \log IC_{50}$. L'avantage principal de cette technique par rapport à l'essai de liaison par saturation est qu'elle requiert une quantité fortement réduite de ligand radioactif, puisqu'il n'est pas requis d'augmenter la concentration de ligand radioactif au-delà du K_d du récepteur. Cela implique des quantités de ligand radioactif de l'ordre du dixième de celles utilisées en essais de saturation, ce qui présente plusieurs avantages en ce qui a trait au coût et à la sécurité de l'expérience.

La principale restriction de cet essai est le signal spécifique maximum :

$$N_{max} - N_{min} = \frac{k_{rad}[R]_{tot}[L^*]}{K_d + [L^*]}$$

Ainsi, dans un contexte où ce type d'essai est utilisé sur les mutants d'un récepteur et que le K_d est altéré (devient ici K'_d), le signal S de liaison spécifique peut être perdu :

$$\begin{aligned} \frac{d}{dt}N'(t) &= S' = \frac{k_{rad}[R']_{tot}[L^*]}{K'_d + [L^*]} \\ \frac{S'}{S} &= R_S = \frac{k_{rad}[R']_{tot}[L^*]}{K'_d + [L^*]} \cdot \frac{K_d + [L^*]}{k_{rad}[R]_{tot}[L^*]} = \frac{[R']_{tot}}{K'_d + [L^*]} \cdot \frac{K_d + [L^*]}{[R]_{tot}} \end{aligned}$$

Dans un contexte où $[R]_{tot}$ et $[R']_{tot}$ diffèrent peu et que $[L^*]$ se situe près de K_d , alors :

$$(12) \quad \frac{k_{rad}[R']_{tot}[L^*]}{K'_d + [L^*]} \cdot \frac{K_d + [L^*]}{k_{rad}[R]_{tot}[L^*]} = R_{signal} = \frac{2K_d}{K'_d + K_d} = \frac{2K_d}{\left(\frac{K'_d}{K_d} + 1\right)K_d} = \frac{2}{\frac{K'_d}{K_d} + 1}$$

Si $K'_d \approx 10K_d$, par exemple :

$$R_{signal} = \frac{2}{11}$$

Cela veut dire que le rapport bruit/signal vient d'à-peu-près quintupler. Au-delà d'une certaine perte d'affinité pour le ligand radioactif, il devient donc impossible d'effectuer un essai de liaison compétitif sans augmenter conséquemment la concentration de ligand radioactif. Comme le gain de radioactivité maximale en fonction de $\log[L^*]$ diminue après $\log K_d$, c'est à une concentration de ligand radioactif similaire à K_d que ces essais sont effectués. Or, dans le contexte de ce travail, le K_d du ligand radioactif pour la totalité des récepteurs mutants est inconnue. Cela signifie que, pour certains mutants, il sera impossible de déterminer les différents paramètres de la liaison à l'apeline en raison d'un K_d trop augmenté. Pour cette raison, et pour confirmer avec une autre technique que le déplacement observé du ligand radiomarqué est bien dû aux récepteurs étudiés exprimés à la surface de la cellule, les différents mutants seront examinés à l'aide d'un ELISA de surface.

Les ELISAs de surface

Les ELISAs de surface sont une forme d'ELISA où les antigènes se trouvent sur la membrane cytoplasmique des cellules. Cette technique consiste en la fixation de cellules exprimant la protéine d'intérêt dans un puits, de leur exposition à un ou plusieurs anticorps en série, dont un spécifique à la protéine d'intérêt et un fixé à une enzyme, puis du retrait des anticorps libres. Les complexes antigène-anticorps sont par la suite exposés à un substrat de l'enzyme attachée au complexe, puis la concentration du produit de l'enzyme est mesurée. Dans le contexte de ce travail, un anticorps anti-HA couplé à la peroxydase du raifort est utilisé (Schönege *et al*, 2017; Murza *et al*, 2016). Les récepteurs utilisés dans ce travail sont tous supplémenté de deux étiquettes HA avant le N-terminus (Murza *et al*, 2016; Brizzard, 2008).

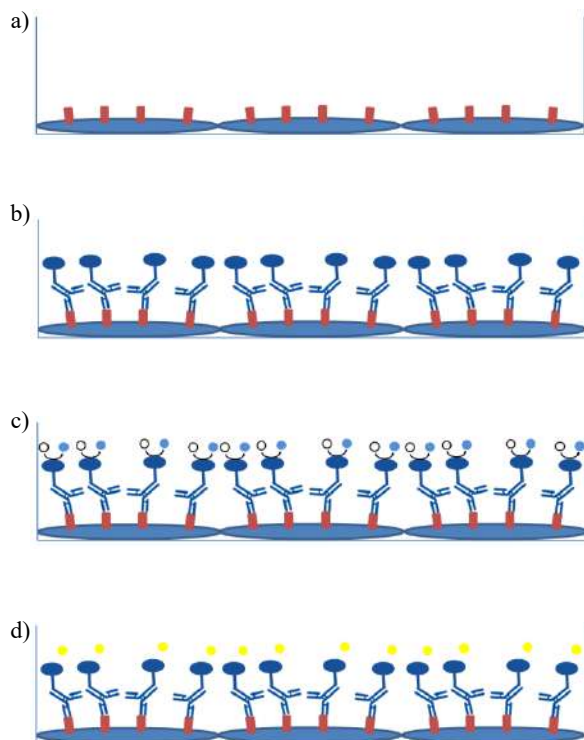


Figure 12 : Description des étapes d'un ELISA de surface

a) D'abord, les cellules sont fixées dans le puits où s'effectue l'analyse. b) Un anticorps contre l'antigène de surface recherché, associé à une enzyme servant à l'étape suivante, est ensuite ajouté. La qualité du résultat dépend de la sensibilité et de la spécificité de cet anticorps. c) Un composé est ajouté (e.g. le TMB) et réagit avec l'enzyme attachée à l'anticorps. La quantité de substrat converti dépend de la quantité d'antigènes présents sur la surface des cellules. d) Finalement, la réaction est stoppée par l'ajout d'un inhibiteur quelconque et l'absorbance de la solution est mesurée à la longueur d'onde appropriée. Ici, il s'agit de HCl, qui stoppe la réaction et donne une couleur jaune au produit de réaction.

Ce type d'essai est utilisé pour quantifier l'internalisation du récepteur et peut aussi être utilisé pour quantifier le rapport d'expression d'un récepteur par rapport à une référence. Ce rapport d'expression peut ensuite être associé à l'intensité de l'activité basale ou maximale d'un récepteur pour différents essais fonctionnels (Schönege *et al*, 2017; Murza *et al*, 2016). Dans ce travail, l'ELISA de surface est surtout utilisé pour confirmer l'expression du récepteur à la surface de la cellule en mesurant la différence de concentration du produit de l'enzyme pour un essai effectué sur le récepteur par rapport à celle d'un essai effectué sur un plasmide n'exprimant que l'étiquette HA sans le récepteur.

Les essais BRET

Les essais BRET sont une sous-catégorie des essais qui exploitent le phénomène de transfert d'énergie par résonance (RET, pour Resonance Energy Transfer). Ce transfert d'énergie survient lorsque deux molécules, un fluorophore « donneur » et un fluorophore « accepteur », se trouvent à proximité l'une de l'autre. Lorsque le fluorophore donneur est excité, une partie de l'énergie qu'il absorbe est transférée de manière non-radiative, décrit

rigoureusement par un photon virtuel (Andrews, 2009), au fluorophore accepteur, qui émet alors un photon à la place du donneur.

Lors du RET, le transfert d'énergie est inversement proportionnel à la distance entre les deux fluorophores :

$$k_t = \frac{1}{\tau_D} \left(\frac{R_0}{r_{DA}} \right)^6$$

r_{DA} représente la distance entre le donneur et l'accepteur et $\frac{1}{\tau_D}$ est le taux de dé-excitation du donneur via toutes les autres voies alternatives au RET. La distance R_0 dépend de l'efficacité de transfert d'énergie entre les deux fluorophores, et vaut (R_0^6 en cm^6) :

$$R_0^6 = 8,79 \cdot 10^{-25} n^{-4} Q_D \kappa^2 \int_0^\infty \varepsilon_A(\bar{\nu}) F_D(\bar{\nu}) \bar{\nu}^{-4} d\bar{\nu} / \int_0^\infty F_D(\bar{\nu}) d\bar{\nu}$$

La valeur des deux intégrales dépend de la bonne superposition des spectres d'absorption de l'accepteur ($\varepsilon_A(\bar{\nu})$) et d'émission de fluorescence du donneur ($F_D(\bar{\nu})$). Q_D est le rendement quantique du donneur, n est l'indice de réfraction du milieu et κ est une mesure de l'orientation relative des deux dipôles :

$$\kappa = \hat{\mathbf{p}}_D \cdot \hat{\mathbf{p}}_A - 3(\hat{\mathbf{r}}_{DA} \cdot \hat{\mathbf{p}}_D)(\hat{\mathbf{r}}_{DA} \cdot \hat{\mathbf{p}}_A)$$

\mathbf{p}_D et \mathbf{p}_A représentent respectivement le dipôle du donneur et celui de l'accepteur et les vecteurs $\hat{\mathbf{v}}$ représentent les vecteurs unitaires correspondants, soit $\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$.

Une explication plus détaillée du phénomène de RET et de ses différentes applications, ainsi que des références additionnelles sur le sujet, peuvent être trouvées dans Clegg, 2008 et Andrews, 2009.

Dans le contexte de ce mémoire, l'application du RET utilise des fluorophores biologiques, nommément pour donneur la coelenterazine, activée par la luciférase de *Renilla reniformis*

(RLucII) et pour accepteur la protéine fluorescente verte, ou GFP, d'*Aequorea victoria* ou une de ses variantes (Pfleger et Eidne, 2006). L'efficacité du RET témoigne ici idéalement de la distance entre deux protéines chimériques, soit par exemple, dans les cas qui nous intéressent, la protéine $G\alpha_i$ -RLucII et la protéine $G\gamma_1$ -GFP10 (Galès *et al*, 2006), ainsi que la protéine β -arrestine 2-RLucII et rGFP précédée du motif CAAX (Quoyer *et al*, 2013; Namkung *et al*, 2016).

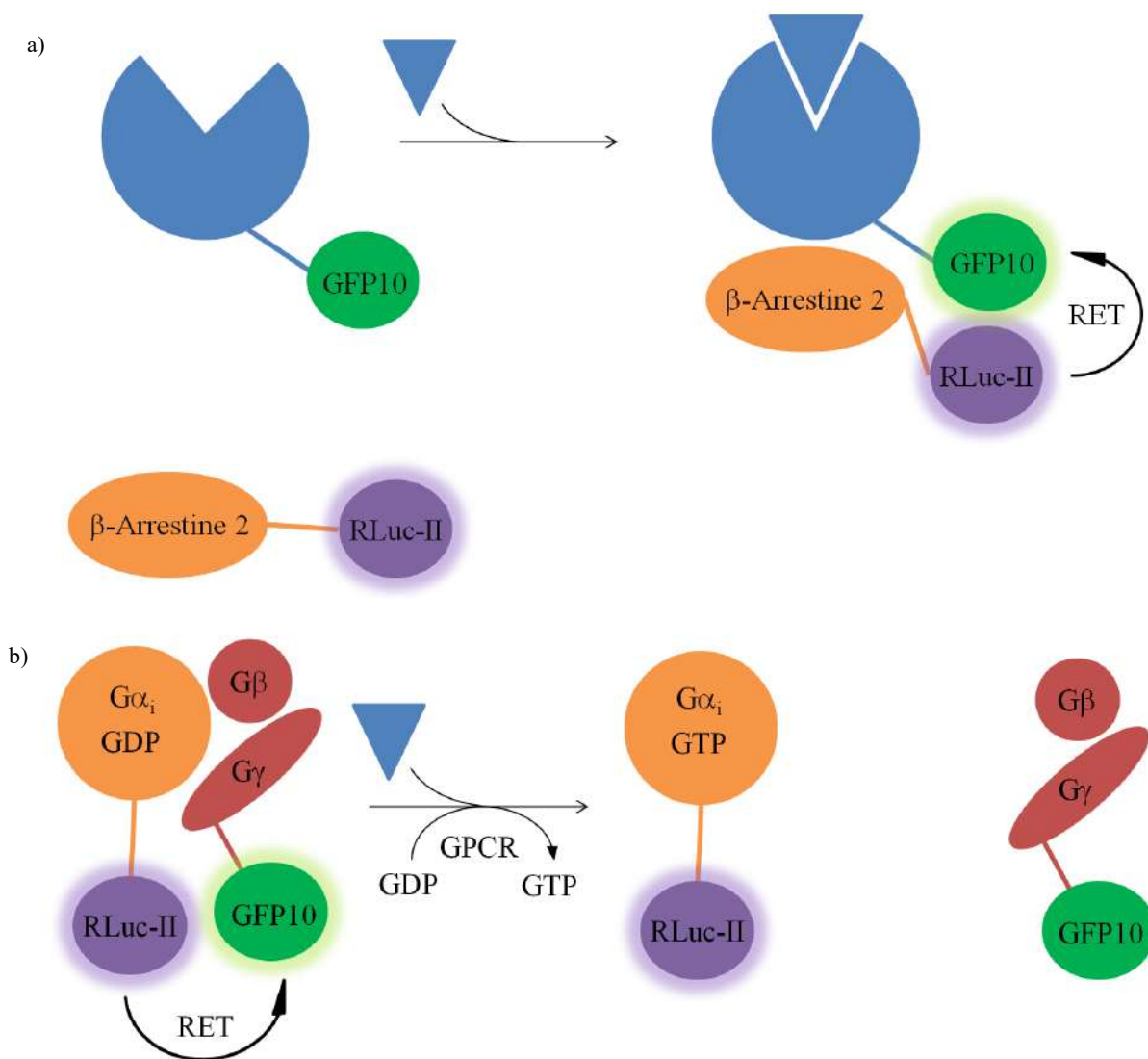


Figure 13 : Fonctionnement du BRET pour les interactions entre un GPCR et les protéines G et les arrestines

a) Le BRET des arrestines s'appuie sur le recrutement des arrestines à la surface de la cellule lorsque les récepteurs sont phosphorylés. Sous l'action du ligand (triangle bleu), le récepteur (cercle bleu) finit par recruter une arrestine. La proximité entre l'arrestine et la membrane cytoplasmique cause le RET. Dans cette figure, la GFP10 est fixée au récepteur, mais dans ce travail, la GFP est fixée à un motif CAAX, qui lui permet de s'ancrer dans la membrane.

b) Le BRET des protéines G se base sur la perte d'affinité entre la protéine $G\alpha$ et le complexe $G\beta\gamma$ lors de son activation. L'activation de la protéine G provoque l'éloignement des deux protéines, éliminant ainsi le RET.

Ceci dit, l'étude de l'interaction de deux protéines par BRET pose un certain risque. En effet, comme les fluorophores sont attachés de façon covalente à des molécules volumineuses qui interagissent via une interface spécifique, leurs degrés de libertés rotationnels sont restreints, ce qui signifie qu'un réarrangement des deux protéines auxquelles les fluorophores sont attachés pourrait causer une observation similaire à un éloignement ou un rapprochement, même s'il n'y en a pas, en raison du changement du paramètre κ . En prenant l'exemple des protéines G, si on considère l'hypothèse que le GPCR forme un pré-complexe avec ces protéines (Ferré, 2015; Navarro *et al*, 2018), alors la comparaison de mutants d'un même récepteur, ou un récepteur exposé à un modulateur allostérique, pose comme risque de comparer des populations de pré-complexes GPCR-complexe $G\alpha\beta\gamma$ chimérique de conformation différente, dont le κ est différent du complexe de référence, donnant ainsi la fausse impression d'une activité basale augmentée ou diminuée. Malgré ce désavantage, le BRET reste une des techniques les plus accessibles pour mesurer la proximité entre deux protéines. De plus, l'utilisation de cette technique est répandue dans le domaine de l'étude fonctionnelle des GPCRs (Schönege *et al*, 2017) et différents protocoles existent déjà et sont optimisés pour APJ (Besserer-Offroy *et al*, 2018; Murza, Sainsily *et al*, 2017; Trân *et al*, 2018). Pour ces raisons, malgré les possibles désavantages associés à la technique, le BRET est utilisé dans ce travail pour étudier la capacité du récepteur à activer la protéine $G\alpha_{i1}$ et recruter la β -arrestine 2.

Dynamique moléculaire des systèmes hAPJ-apeline et hAPJ

Construction des systèmes hAPJ-apeline et hAPJ

Pour générer ces systèmes, la structure du récepteur APJ en complexe avec un ligand agoniste peptidique macrocyclique a été utilisée (Ma *et al*, 2017). La structure d'APJ contenait plusieurs mutations qui ont été corrigées en utilisant Modeller (Webb et Sali, 2014) avec la véritable structure primaire d'APJ humain (hAPJ) comme structure primaire à modéliser et la structure d'APJ comme modèle. Considérant le taux d'homologie élevé des deux structures (plus de 95%), il est raisonnable de croire que le modèle est aussi fidèle que possible. Considérant que les résidus de la portion N-terminale précédant la cystéine 19, ainsi que les résidus de la portion C-terminale de la protéine sont fortement mobiles et

que leur comportement dynamique est mal connu, ils sont retirés du modèle. De même, les résidus de l'ICL3 n'étant pas résolus dans la structure d'APJ disponible, cette portion d'APJ a été reconstituée de façon à ce que sa structure interfère le moins possible avec la reste de la protéine. Étant donné la présence importante de résidus basiques dans cette région, plus particulièrement d'arginines, cette région de la protéine est ancrée dans la membrane lors des simulations.

L'apeline a été générée directement à partir du ligand peptidique. En effet, il s'agit de la structure la plus proche qu'il est actuellement possible d'obtenir d'un ligand peptidique agoniste d'APJ. Ainsi, le ligand a été muté en apeline-13 à l'aide de PyMOL (Schrödinger LLC).

Une fois les modèles de hAPJ et de l'apeline obtenus, il a suffi de combiner de nouveau les deux structures pour obtenir le complexe hAPJ-apeline. La structure de hAPJ seul et du complexe ont ensuite été plongés dans une membrane de DOPC/Cholestérol (Wennberg *et al*, 2012) entourée d'eau (modèle SPC) et équilibrés à 310K et 1 atm à l'aide de la suite GROMACS (Abraham *et al*, 2015; Berendsen *et al*, 1995).

Dynamique moléculaire de production

Une fois les systèmes construits et équilibrés, les simulations de dynamique moléculaire pour les deux systèmes ont été effectuées tel que décrit précédemment. Les huit simulations en parallèle par round de simulation ont été effectuées sur le mammoth-mp2 de l'Université de Sherbrooke. Une fois les résultats bruts obtenus, les angles dièdres ont été calculés en utilisant la librairie MDTraj (McGibbon *et al*, 2015). Comme certaines structures de chaînes latérales transformées par la rotation de certains angles dièdres symétrie C_2 , nommément le χ_2 de la phénylalanine, de la tyrosine et de l'aspartate, ainsi que le χ_3 du glutamate, c'est leur $\text{Sin}^2\theta$ qui a été calculé au lieu de leur $\text{Sin}\theta$, car sa période est réduite de moitié, ce qui illustre mieux le changement de ces angles dièdres.

Les résultats cumulatifs analysés avec la combinaison tICA/GMHMM. Les changements conformationnels responsables de la séparation des états du GMHMM ont été identifiés en utilisant la distance de Bhattacharyya et les résidus les plus impliqués dans ces

changements conformationnels ont été identifiés par une analyse des sous-matrices principales. Ces deux analyses sont décrites avec plus de détails dans la prochaine section.

Analyse des simulations de dynamique moléculaire : TOCATA-MD

Simplification des données : tICA/GMHMM

L'analyse d'un seul ensemble de simulations de dynamique moléculaire est un processus relativement trivial pour les protéines lorsqu'on utilise la combinaison tICA/GMHMM. En effet, des bibliothèques existent déjà pour extraire et analyser toutes les informations voulues, nommément MDTraj (McGibbon *et al*, 2015) et MSMBuilder (Beauchamp *et al*, 2011). Dans le contexte de ce travail, une partie du défi réside plutôt dans la rapidité et l'optimalité de l'analyse, ainsi que dans l'interprétation de l'analyse.

En effet, le nombre de systèmes à analyser dans un contexte de ré-échantillonnage des états minoritaires, et considérant le nombre de conditions différentes du système hAPJ à analyser, s'élève à une cinquantaine. Pour chacune de ces analyses, il faudrait faire varier manuellement le nombre de dimensions et d'états jusqu'à trouver le modèle optimal. Étant donné le temps énorme que cela prendrait et les contraintes de temps serrées du projet, une automatisation et une optimisation de l'analyse est nécessaire.

Pour ce faire, un programme a été construit selon les spécifications données dans l'introduction :

1. Transformer les données brutes d'angles dièdres en leur sinus ou cosinus
2. Projeter les données conformationnelles brutes sur le premier vecteur propre de haute autocorrélation obtenus par tICA.
3. Grouper les données et construire une chaîne de Markov en optimisant un GMHMM.
4. Répéter 3 en ajoutant un état au système jusqu'à optimisation du BIC.
5. Répéter 2, 3 et 4 en ajoutant un vecteur propre de la tICA à chaque itération, jusqu'à optimisation de la seconde échelle de temps de la chaîne de Markov issue du système.

Deux observations supplémentaires sont prises en compte :

- En pratique, le BIC suit une décroissance exponentielle, suivi d'un long plateau autour du minimum, puis d'une croissance graduelle, mais faible. Ainsi, il est possible de choisir un modèle marginalement moins bon que vrai modèle optimal, mais beaucoup plus rapidement et avec beaucoup moins d'états.
- Toujours en pratique, l'échelle de temps t_2 du GMHMM semble plafonner rapidement et n'augmente plus après un certain nombre de dimensions. Encore une fois, il est possible de sélectionner un modèle marginalement moins bon que le modèle optimal, mais en beaucoup moins de temps

Il faut se rappeler que l'objectif principal n'est pas absolument d'obtenir un modèle optimal. L'idée est d'obtenir un modèle informatif duquel il est possible de tirer un maximum d'information en un temps raisonnable. Or, l'exploration de l'espace des états du GMHMM s'effectue en un temps quadratique ($O(n^2)$) pour un nombre d'états donné, ce qui signifie que l'exploration du nombre d'états est un processus en temps cubique, car $\sum_{i=1}^n i^2 \in O(n^3)$. C'est sans compter le fait que ce temps n'est que pour un ensemble de composantes de la tICA et qu'il y a possiblement une dizaine d'ensembles de composantes à tester dans le pire des cas. Éliminer quelques étapes donnant un gain marginal en qualité sauve donc un temps énorme dans ces conditions.

L'implémentation de l'optimisation est donc la suivante : chaque ajout d'état ou de dimension doit donner au minimum une amélioration de 5% du BIC ou de l'échelle de temps par rapport au modèle minimal (i.e. un seul état pour le BIC et une dimension de moins pour l'échelle de temps. Cette amélioration, afin d'éviter un arrêt trop rapide, peut être étendue jusqu'à l'ajout de trois états pour le BIC et deux dimensions pour l'échelle de temps, i.e. si $i + 3$ états donnent une amélioration de 5% par rapport à i états, le modèle est conservé. Cet assouplissement du critère d'amélioration tient en compte que la valeur du BIC est elle-même une variable aléatoire, découlant de la nature non-déterministe des paramètres initiaux des états, ainsi que des paramètres choisis pour le processus d'optimisation, ce qui résulte en une incertitude sur le BIC. Il est possible que le BIC soit significativement plus haut que sa moyenne, ce qui donnerait un signal d'arrêt d'optimisation alors que le plateau n'est pas encore vraiment atteint. Une succession de

trois erreurs positives importantes sur le BIC est plus rare et constitue donc une certaine forme de sécurité à un coût relativement faible. De plus, comme le signal d'arrêt d'optimisation encourage une analyse moins détaillée du système, cette sécurité permet de compenser un peu ce défaut. Le même raisonnement s'applique pour l'échelle de temps, qui est optimisée par le même processus.

Analyse des données : sous-matrices principales

Une fois les données brutes simplifiées en leurs composantes de tICA et états GMHMM, une question subsiste : que représentent réellement ces composantes et états, et surtout, est-il possible d'utiliser notre compréhension de la nature de ces résultats pour les rendre encore plus informatifs?

D'abord, il serait intéressant de voir quelles informations supplémentaires il serait possible d'obtenir des résultats de la tICA. Pour ce faire, il est utile de considérer la nature des vecteurs propres généralisés. Le i -ième vecteur propre généralisé de l'équation $C_s(\tau)\mathbf{v}_i = \lambda_i \Sigma \mathbf{v}_i$ (forme à un vecteur de l'équation 7) correspond également à la combinaison linéaire des composantes brutes (e.g. angles dièdres) du système qui maximise l'autocorrélation tout en étant Σ -orthogonal aux $i - 1$ vecteurs propres précédents (i.e. $\mathbf{v}_i^T \Sigma \mathbf{v}_j = \delta_{ij}$). La preuve de cette affirmation est similaire à celle du principe variationnel des processus markoviens, les deux étant des problèmes d'optimisation utilisant le théorème spectral comme avenue de résolution. Or, ce que le vecteur propre ne dit pas, c'est à quel point chaque composante brute est importante pour cette maximisation. Par exemple, une composante brute pourrait amener à elle seule l'autocorrélation à 90% de son maximum, mais une combinaison linéaire de 200 autres composantes brutes serait nécessaire pour l'emmener à 100% de son maximum. Une telle situation est peu probable, mais l'argument est similaire même si une poignée de composantes brutes est responsable de la majorité de l'optimisation. La question est donc : « Comment identifier quelles composantes sont responsables de la plus grande part de l'optimisation? »

L'avenue choisie lors de ce travail utilise une variante du théorème de séparation de Poincaré adapté au problème des vecteurs et valeurs propres généralisé (Thompson, 1972).

Une autre solution potentielle à ce problème est présentée dans la discussion. Du système $AV = BV\Lambda$ avec $A, B \in \mathbb{R}^{n \times n}$, considérons le sous-système $A_{-\mathbb{I}}V' = B_{-\mathbb{I}}V'M$, où \mathbb{I} est un ensemble d'indices entre 1 et n et $A_{-\mathbb{I}}$ et $B_{-\mathbb{I}}$ les matrices A et B dont on a retiré les lignes et colonnes dont les indices sont éléments de \mathbb{I} (i.e. $A_{-\mathbb{I}}, B_{-\mathbb{I}} \in \mathbb{R}^{(n-|\mathbb{I}|) \times (n-|\mathbb{I}|)}$). Les matrices $A_{-\mathbb{I}}$ et $B_{-\mathbb{I}}$ sont des sous-matrices principales de A et B . Soient $\lambda_1 > \dots > \lambda_n$ les valeurs propres du système $AV = BV\Lambda$ et $\mu_1 > \dots > \mu_{n-|\mathbb{I}|}$ les valeurs propres de $A_{-\mathbb{I}}V' = B_{-\mathbb{I}}V'M$. Le théorème de séparation de Poincaré stipule que $\lambda_i \geq \mu_i \geq \lambda_{i+|\mathbb{I}|}$.

La justification biophysique de cette analyse est la suivante : soit une série temporelle stochastique de n variables aléatoires (le système étudié). Soit C_τ la matrice d'autocorrélation du système de délai τ et Σ la matrice des covariances du système. Dans un contexte biophysique, si les composantes du système ne contiennent que des coordonnées internes au système (i.e. les degrés de liberté translationnels et rotationnels sont ignorés), alors la matrice Σ est symétrique, définie positive et donc inversible (i.e. la matrice décrit l'approximation du système par une loi normale multivariée). Soit ensuite le système de vecteurs et valeur propres (ici appelé eigensystème par simplicité) suivant :

$$C_\tau \mathbf{v} = \lambda \Sigma \mathbf{v} \Rightarrow (\Sigma^{-1} C_\tau - \lambda I) \mathbf{v} = \mathbf{0}$$

Le polynôme caractéristique $\det(\Sigma^{-1} C_\tau - \lambda I) = 0$ contient des valeurs entre -1 et 1 et représentent les différentes autocorrélations des différents vecteurs propres du système. Une autocorrélation supérieure à zéro signifie que les différentes composantes du système étudié ont une dépendance linéaire entre elles-mêmes via un délai temporel τ . L'objectif de l'analyse par sous-matrices principales est d'étudier l'effet sur le polynôme caractéristique du système (i.e. ses autocorrélations au délai τ) du retrait de cette dépendance linéaire pour une composante par rapport au reste du système. Pour faire le lien entre les deux concepts, considérons le système suivant :

$$C_\tau^- \mathbf{v}^- = \mu \Sigma^- \mathbf{v}^- \Rightarrow ((\Sigma^-)^{-1} C_\tau^- - \mu I) \mathbf{v}^- = \mathbf{0}$$

Ce système contient les même valeurs que $C_\tau \mathbf{v} = \lambda \Sigma \mathbf{v}$, à l'exception d'une ligne et d'une colonne intersectant sur la diagonale, qui ont été retirées. Les matrices C_τ^- et Σ^- sont donc

des sous-matrices principales de C_τ et Σ , respectivement. Soient maintenant les matrices suivantes :

$$C_{\tau,-k} = \begin{bmatrix} c_{11\tau} & c_{12\tau} & \dots & c_{1(k-1)\tau} & 0 & c_{1(k+1)\tau} & \dots & c_{1n\tau} \\ c_{21\tau} & c_{22\tau} & \dots & c_{2(k-1)\tau} & 0 & c_{2(k+1)\tau} & \dots & c_{2n\tau} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ c_{(k-1)1\tau} & c_{(k-1)2\tau} & \dots & c_{(k-1)(k-1)\tau} & 0 & c_{(k-1)(k+1)\tau} & \dots & c_{(k-1)n\tau} \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ c_{(k+1)1\tau} & c_{(k+1)2\tau} & \dots & c_{(k+1)(k-1)\tau} & 0 & c_{(k+1)(k+1)\tau} & \dots & c_{(k+1)n\tau} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n1\tau} & c_{n2\tau} & \dots & c_{n(k-1)\tau} & 0 & c_{n(k+1)\tau} & \dots & c_{nn\tau} \end{bmatrix}$$

$$\Sigma_{-k} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1(k-1)} & 0 & \sigma_{1(k+1)} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2(k-1)} & 0 & \sigma_{2(k+1)} & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ \sigma_{(k-1)1} & \sigma_{(k-1)2} & \dots & \sigma_{(k-1)(k-1)} & 0 & \sigma_{(k-1)(k+1)} & \dots & \sigma_{(k-1)n} \\ 0 & 0 & \dots & 0 & \sigma_{kk} & 0 & \dots & 0 \\ \sigma_{(k+1)1} & \sigma_{(k+1)2} & \dots & \sigma_{(k+1)(k-1)} & 0 & \sigma_{(k+1)(k+1)} & \dots & \sigma_{(k+1)n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{n(k-1)} & 0 & \sigma_{n(k+1)} & \dots & \sigma_{nn} \end{bmatrix}$$

Ces matrices correspondent aux matrices d'autocorrélation et des covariances, respectivement, du système auquel on a substitué la variable aléatoire k par un bruit gaussien de même variance. On remarque notamment que la matrice Σ_{-k} reste définie positive, puisque $\sigma_{kk} > 0$ et que la sous-matrice de Σ_{-k} dont la ligne et la colonne k ont été retirées satisfait aux mêmes propriétés que Σ et est donc définie positive. L'objectif est de trouver une façon d'évaluer l'effet de cette substitution sur le polynôme caractéristique du système, et il advient que :

$$(13) \quad \det(\Sigma_{-k}^{-1} C_{\tau,-k} - \mu I) = -\mu \det((\Sigma^-)^{-1} C_\tau^- - \mu I)$$

En effet, il suffit de considérer le problème à l'envers : connaissant $(\Sigma^-)^{-1}$, il est trivial de démontrer que l'inverse de Σ_{-k} n'est nul autre que $(\Sigma^-)^{-1}$ à laquelle on a ajouté une ligne et une colonne remplies de zéros qui s'intersectent à la position (k, k) , sauf sur la diagonale où la valeur est $\frac{1}{\sigma_{kk}}$. Considérant que tous les éléments hors-diagonale de cette ligne et cette colonne sont zéro, multiplier cette matrice – appelons-la Σ_{-k}^{-1} – avec Σ_{-k} revient à multiplier $(\Sigma^-)^{-1}$ avec Σ^- , sauf pour la paire ligne-colonne ajoutée. Dans le cas de cette

paire, de par sa nature, tout élément hors-diagonale de cette ligne et cette colonne du produit vaut zéro. Sur la diagonale, le calcul se réduit à $\frac{\sigma_{kk}}{\sigma_{kk}} = 1$, ce qui donne en somme la matrice identité.

Ensuite, on montre que $\Sigma_{-k}^{-1} C_{\tau, -k} \sim (\Sigma^-)^{-1} C_{\tau}^-$: multiplier Σ_{-k}^{-1} avec $C_{\tau, -k}$ revient encore une fois, pour la même raison que Σ_{-k} et Σ^- , à multiplier $(\Sigma^-)^{-1}$ et C_{τ}^- , à l'exception de la ligne et la colonne ajoutés qui valent zéro partout :

$$\Sigma_{-k} C_{\tau, -k}$$

$$= \begin{bmatrix} ((\Sigma^-)^{-1} C_{\tau}^-)_{11} & ((\Sigma^-)^{-1} C_{\tau}^-)_{12} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{1(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{1k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{1n} \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{21} & ((\Sigma^-)^{-1} C_{\tau}^-)_{22} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{2(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{2k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)n} \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{k1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{k2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{k(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{kk} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{kn} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{n1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{n2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{n(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{nk} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{nn} \end{bmatrix}$$

Donc nécessairement :

$$\det(\Sigma_{-k}^{-1} C_{\tau, -k} - \mu I)$$

$$= \begin{vmatrix} ((\Sigma^-)^{-1} C_{\tau}^-)_{11} - \mu & ((\Sigma^-)^{-1} C_{\tau}^-)_{12} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{1(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{1k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{1n} \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{21} & ((\Sigma^-)^{-1} C_{\tau}^-)_{22} - \mu & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{2(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{2k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)(k-1)} - \mu & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)n} \\ 0 & 0 & \cdots & 0 & -\mu & 0 & \cdots & 0 \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{k1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{k2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{k(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{kk} - \mu & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{kn} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{n1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{n2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{n(k-1)} & 0 & ((\Sigma^-)^{-1} C_{\tau}^-)_{nk} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{nn} - \mu \end{vmatrix}$$

$$= -\mu \begin{vmatrix} ((\Sigma^-)^{-1} C_{\tau}^-)_{11} - \mu & ((\Sigma^-)^{-1} C_{\tau}^-)_{12} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{1(k-1)} & ((\Sigma^-)^{-1} C_{\tau}^-)_{1k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{1n} \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{21} & ((\Sigma^-)^{-1} C_{\tau}^-)_{22} - \mu & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{2(k-1)} & ((\Sigma^-)^{-1} C_{\tau}^-)_{2k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)(k-1)} - \mu & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)k} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{(k-1)n} \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{k1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{k2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{k(k-1)} & ((\Sigma^-)^{-1} C_{\tau}^-)_{kk} - \mu & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{kn} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ ((\Sigma^-)^{-1} C_{\tau}^-)_{n1} & ((\Sigma^-)^{-1} C_{\tau}^-)_{n2} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{n(k-1)} & ((\Sigma^-)^{-1} C_{\tau}^-)_{nk} & \cdots & ((\Sigma^-)^{-1} C_{\tau}^-)_{nn} - \mu \end{vmatrix}$$

$$= -\mu \det((\Sigma^-)^{-1} C_{\tau}^- - \mu I) \blacksquare$$

On sait par conséquent que l'étude du retrait de la dépendance linéaire d'une composante d'un système avec les autres via un certain délai sur l'autocorrélation peut passer par l'étude des sous-matrices principales correspondantes. De même, comme les matrices Σ^- et C_{τ}^- ont les mêmes propriétés que Σ et C_{τ} , on peut choisir d'étudier le retrait de cette

dépendance pour un groupe de variables en même temps. Par exemple, on peut autant étudier l'effet du retrait de la dépendance d'un seul angle dièdre d'un résidu que celui de tous les angles dièdres d'un résidu. De là, il est intéressant de s'interroger sur l'amplitude de cette réduction pour chaque résidu qu'on rend indépendant du reste du système.

Pour un résidu r , soit $C_{S(-\mathbb{I}_r)}(\tau)\mathbf{v}'_i = \mu_i \Sigma_{-\mathbb{I}_r} \mathbf{v}'_i$ l'équation pour le i -ième vecteur propre de ce système modifié, avec \mathbb{I}_r l'ensemble des indices correspondant à ces résidus. Par le théorème de séparation de Poincaré, on sait que $\lambda_i \geq \mu_i \geq \lambda_{i+|\mathbb{I}_r|}$ ($|\mathbb{I}_r|$ représente ici la cardinalité de l'ensemble). Soit donc la quantité $\rho_{r,i} = \frac{\mu_i - \lambda_{i+|\mathbb{I}_r|}}{\lambda_i - \lambda_{i+|\mathbb{I}_r|}}$. Cette quantité a été utilisée pour quantifier la diminution des valeurs propres du système réduit. Pour s'assurer que la comparaison des valeurs propres est bien la bonne, le $\cos\theta$ entre les vecteurs propres \mathbf{v}_i , moins les composantes brutes qui ont été retirées, et les vecteurs propres \mathbf{v}'_i ont aussi été calculés.

Analyse des données : distance de Bhattacharyya

Ensuite, une fois le rôle de chaque résidu élucidé pour chaque paire de vecteur propre-valeur propre de la tICA, d'autres informations pertinentes peuvent être extraites du système, cette fois du GHMM. L'information la plus intéressante dans le cadre de ce travail est la séparation du système sous-ensembles conformationnels, en déterminant simultanément quelles composantes indépendantes sont responsables de la séparation de ces états.

Pour ce faire, comme le système est désormais réduit à un ensemble de Gaussiennes dont la matrice des covariances est diagonale, une solution directe au problème des séparations d'ensembles conformationnels est identifiable en utilisant la distance de Bhattacharyya (Bhattacharyya, 1943; Choi et Lee, 2003). Soient les variables aléatoires $X \sim \mathcal{N}(\boldsymbol{\mu}_1; \Sigma_1)$ et $Y \sim \mathcal{N}(\boldsymbol{\mu}_2; \Sigma_2)$. La distance de Bhattacharyya entre ces deux variables vaut :

$$D_B(X, Y) = \frac{1}{8} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \left(\frac{\Sigma_2 + \Sigma_1}{2} \right)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \frac{1}{2} \ln \left(\frac{|\frac{\Sigma_2 + \Sigma_1}{2}|}{\sqrt{|\Sigma_2| |\Sigma_1|}} \right)$$

Or, les matrices des covariances sont diagonales. Cette expression peut donc être réduite :

$$\begin{aligned} D_B(X, Y) &= \frac{1}{8} \sum_{i=1}^n \left(\frac{2}{\Sigma_{2(i,i)} + \Sigma_{1(i,i)}} \right) (\mu_{2,i} - \mu_{1,i})^2 + \frac{1}{2} \ln \left(\frac{\prod_{i=1}^n \left(\frac{\Sigma_{2(i,i)} + \Sigma_{1(i,i)}}{2} \right)}{\sqrt{\prod_{i=1}^n \Sigma_{2(i,i)} \prod_{i=1}^n \Sigma_{1(i,i)}}} \right) \\ &= \frac{1}{8} \sum_{i=1}^n \frac{(\mu_{2,i} - \mu_{1,i})^2}{\Sigma_{12(i,i)}} + \frac{1}{2} \ln \left(\prod_{i=1}^n \left(\frac{\Sigma_{12(i,i)}}{\sqrt{\Sigma_{2(i,i)} \Sigma_{1(i,i)}}} \right) \right), \Sigma_{12(i,i)} = \frac{\Sigma_{2(i,i)} + \Sigma_{1(i,i)}}{2} \\ &= \frac{1}{8} \sum_{i=1}^n \frac{(\mu_{2,i} - \mu_{1,i})^2}{\Sigma_{12(i,i)}} + \frac{1}{2} \sum_{i=1}^n \ln \left(\frac{\Sigma_{12(i,i)}}{\sqrt{\Sigma_{2(i,i)} \Sigma_{1(i,i)}}} \right) \\ (14) \quad D_B(X, Y) &= \sum_{i=1}^n \left(\frac{1}{8} \frac{(\mu_{2,i} - \mu_{1,i})^2}{\Sigma_{12(i,i)}} + \frac{1}{2} \ln \left(\frac{\Sigma_{12(i,i)}}{\sqrt{\Sigma_{2(i,i)} \Sigma_{1(i,i)}}} \right) \right) \end{aligned}$$

Les termes individuels de cette somme peuvent être interprétés comme la contribution de chaque composante indépendante à la distance de Bhattacharyya. Cette distance est associée à l'erreur de classification d'observations issues des deux distributions. Dans le contexte d'un GHMM, cela signifie que les ensembles observations dont les états associés ont une grande distance entre eux sont faciles à distinguer l'un de l'autre. Dans le contexte plus spécifique de contribution d'une composante à cette distance, une contribution élevée signifie que considérer seulement cette seule composante est suffisant pour qu'on observe la séparation des états. Autrement dit, cette composante serait responsable de la séparation conformationnelle de ces états. Une valeur de distance de 1,982 correspond à une erreur de classification moyenne de 2,5% (Choi et Lee, 2003). C'est cette valeur qui est utilisée pour la classification.

Gestion semi-automatique des simulations en série : fonctionnalités de TOCATA-MD

L'ensemble ces techniques supplémentaires, ainsi que des fonctionnalités dédiées à l'optimisation du processus de simulation en série, a ensuite été intégré dans un programme dont le nom préliminaire est TOCATA-MD, pour (Time-Oriented, Conformation-Agregating Tool for the Analysis of Molecular Dynamics). Ce programme s'appuie sur la suite MDTraj (McGibbon *et al*, 2015) et MSMBuilder (Beauchamp *et al*, 2011) et y ajoute les fonctionnalités d'automatisation de sélection de modèle demandées par ce travail :

- Sélection personnalisée des paramètres à analyser de façon plus flexible que par les sélections via la topologie MDTraj,
- Sélection automatique d'un modèle GHMM optimal,
- Impression des résultats de base de l'analyse et des commandes à entrer pour démarrer les prochaines simulations.

Ainsi, l'effort de l'utilisateur se limite à entrer les paramètres de l'analyse et utiliser les résultats pour démarrer les prochaines simulations. Les dont il vient d'être question ont été également incluses dans TOCATA-MD afin de permettre une analyse plus poussée du système au besoin. Le code Python du programme est disponible à l'adresse https://bitbucket.org/lbcossette/tocata_md/src/remake/.

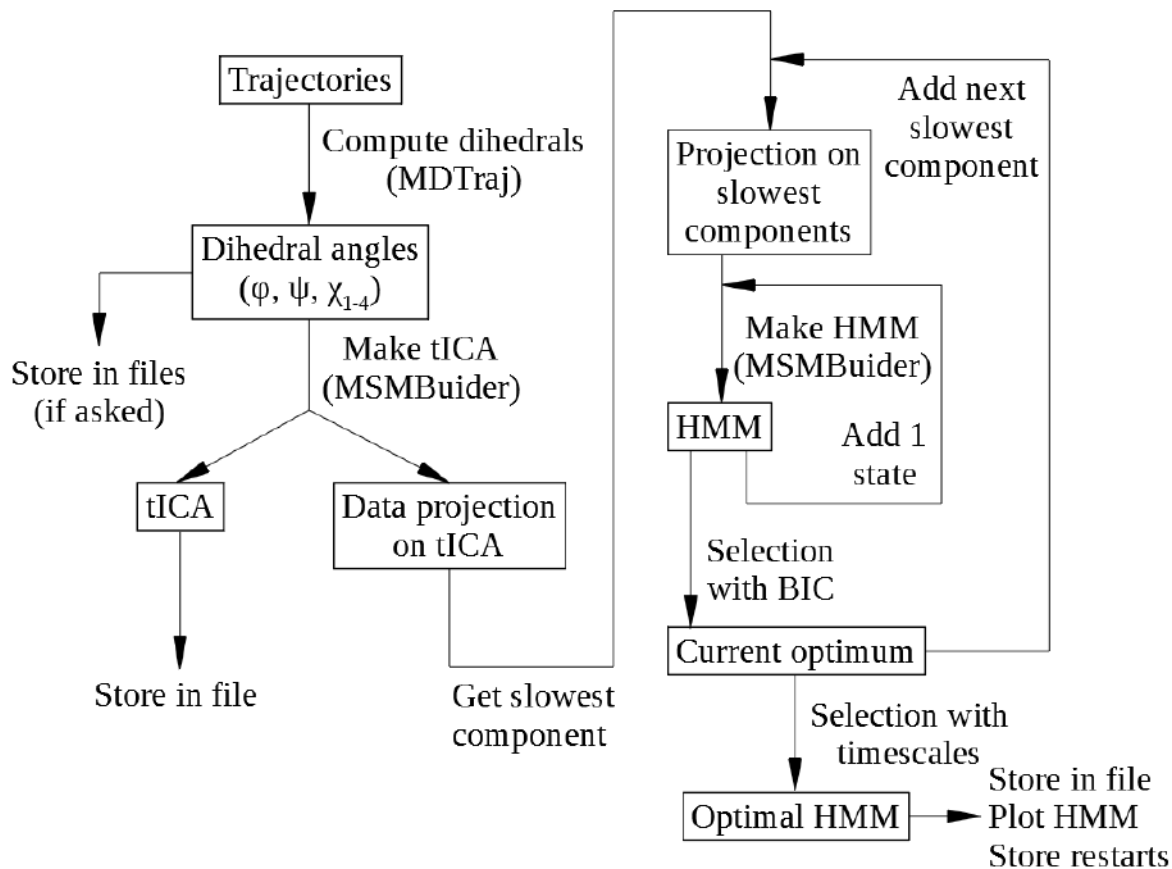


Figure 14 : Description du protocole de base de TOCATA-MD

L'information des trajectoires est d'abord transformée de positions à angles dièdres, la succession des images est analysée avec la tICA, puis projetée sur les composantes les plus lentes. Ensuite, l'ensemble des projections, en commençant par la dimension correspondant à la composante la plus lente, est analysé avec un GMHMM, en ajoutant un état par itération jusqu'à optimisation du BIC. Ensuite, une dimension de la projection est ajoutée et le processus d'optimisation du BIC répété. Le HMM optimal actuel est comparé au nouveau HMM avec leurs échelles de temps respectives, puis le nouveau HMM optimal est choisi. Une fois le processus terminé, le modèle optimal est stocké pour usage ultérieur, des graphiques du modèle sont produits et les états pouvant être utilisés pour démarrer un prochain round de simulations sont produits.

Génération des mutants de hAPJ

Matériel de mutagenèse par QuickChange, transformation bactérienne et extraction d'ADN

- Plasmide d'APJ WT issu d'un pYFPC, dont la YFP a été changée pour hAPJ muni d'une double étiquette HA en N-terminus (Murza, Sainsily *et al*, 2017) (sélection par kanamycine). Cette chimère sera appelée 2HA-hAPJ.
- Plasmide « Mock » issu d'un pYFPC, dont la YFP a été changée pour une double étiquette HA (sélection par kanamycine)
- Plasmide codant pour $G\alpha_{i1}$ -RLucII (Galés *et al*, 2006) (sélection par carbénicilline)

- Plasmide codant pour G β 1, obtenu via cDNA.org (sélection par carbénicilline)
- Plasmide codant pour G γ 1-GFP10 (Galés *et al*, 2006) (sélection par carbénicilline)
- Plasmide codant pour β -arrestine 2-RLucII (Quoyer *et al*, 2013) (sélection par carbénicilline)
- Plasmide codant pour CAAX-rGFP (Namkung *et al*, 2016) (sélection par carbénicilline)
- ADN polymérase Pfu Turbo Cx Hotstart avec tampon dédié et DMSO (Agilent 600410)
- Mélange de dNTPs 100mM (Agilent 200415)
- Enzyme de restriction DpnI (New England Biolabs R0176S)
- Machine de PCR (Bio-Rad T100 Thermal Cycler)
- Tryptone (BD Biosciences 211705)
- Extrait de levure autolysée (BD Biosciences 212750)
- NaCl (EMD-Millipore SX0420-5)
- Sulfate de kanamycine (Thermo Fisher Scientific 11815024)
- Carbénicilline (ThermoFisher Scientific 10177012)
- Agar (Difco (Fisher Scientific) 0145-17-0)
- MiniPrep d'ADN plasmidique (BioBasic 9k-006-0010 and Omega Bio-Tek CA101318-896)
- MidiPrep d'ADN plasmidique (Omega Bio-Tek D6904-03)
- MaxiPrep d'ADN plasmidique (Omega Bio-Tek D6922-01)

Méthode de mutagenèse par QuickChange

La mutagenèse par QuickChange est basée sur la méthode de réaction en chaîne avec polymérase (PCR) (Ho *et al*, 1989).

Dans un tube de PCR à paroi fine, les ingrédients suivants sont mélangés pour chaque réaction :

- 30 ng de plasmide de 2HA-hAPJ,
- 125 ng de l'amorce 5' appropriée (voir Annexe I),
- 125 ng de l'amorce 3' appropriée (voir Annexe I),
- 5 μ L de tampon Pfu 10X,

- 0,4 μL de solution de dNTP,
- Approximativement 2,5 μL de DMSO pour les mutagenèses plus difficiles,
- Le tout complété à 50 μL avec de l'eau dé-ionisée.

Ensuite, au moment de démarrer la réaction, 1 μL (2,5 U) de Pfu Turbo Cx Hotstart est ajouté et mélangé au reste de la solution. Le mélange est ensuite immédiatement inséré dans la machine à PCR, qui est démarrée avec les paramètres de cycle suivants :

- A. 95°C pour 30 s
- B. 95°C pour 20 s
- C. 55°C pour 60 s
- D. 68°C pour 1 min par kilo-paire de base, arrondi à l'entier supérieur, plus un
- E. Go to B (18 cycles)
- F. 4°C (fin de la réaction)

Une fois le mélange récupéré, l'enzyme DpnI est ajoutée et le mélange incubé pendant 60 minutes à 37°C.

Méthode de transformation bactérienne et de purification des plasmides

À 50 μL de culture bactérienne d'*Escherichia coli* DH5 α perméabilisées au CaCl_2 , sur glace, sont ajoutés 2 μL de produit de PCR, ou 50 ng de plasmide intact. Le mélange est agité doucement et incubé sur glace pendant 25 minutes, suite à quoi il est porté à 42°C pendant 50 s, puis ramené sur glace pendant 2 min. 300 μL de milieu LB (1% (m/v) tryptone, 0,5% (m/v) extrait de levure autolysée, 1% (m/v) NaCl) est ensuite ajouté et le mélange est incubé 60 min à 37°C. Pendant ce temps, des pétris contenant du milieu LB gélifié avec de l'agar sont réchauffés 30 min à 37°C. Une solution de carbénicilline ou de sulfate de kanamycine, selon le plasmide utilisé, est ensuite étendue sur la surface du pétri, de façon à obtenir une concentration de 100 $\mu\text{g/mL}$ pour la carbénicilline et 50 $\mu\text{g/mL}$. Le plasmide est ensuite gardé à 37°C jusqu'à ce que les bactéries aient terminé leur incubation. Ces bactéries sont étendues à raison de 350 μL par pétri pour les transformations avec produits de PCR et 100 μL par pétri pour les transformations avec plasmides intacts. Les pétris sont ensuite incubés pendant une nuit (12 à 20 h) jusqu'à l'apparition de colonies bactériennes clonales.

Une fois les colonies développées, un échantillon des colonies (une seule colonie pour les plasmides intacts et jusqu'à trois pour les produits de PCR, selon la quantité disponible) est prélevé et incubé pendant une nuit à 37°C dans un milieu LB supplémenté de l'antibiotique approprié aux mêmes concentrations que pour la transformation bactérienne. Le volume de culture bactérienne utilisée varie selon le kit utilisé. Pour les MiniPreps, le volume est de 6 mL, pour les MidiPreps, de 75 mL et pour les MaxiPreps, de 200 mL. Pour les MidiPreps et les MaxiPreps, une pré-incubation des bactéries dans 6 mL du même milieu pendant 6 à 8h est effectuée afin de mieux contrôler et accélérer le processus. Le reste des instructions est conforme à celles du kit utilisé.

Une fois les produits de PCR purifiés, les plasmides résultants sont envoyés pour validation à la plate-forme de séquençage et de génotypage des génomes de l'université Laval. Si leur mutation est validée, ils sont utilisés pour les expériences subséquentes.

Analyse expérimentale des mutants de hAPJ

Matériel générique pour la culture cellulaire et les expériences associées

- Puits de culture cellulaire de 10 cm (Sarstedt 83.3902)
- Plaque de 6 puits de culture cellulaire (Falcon (Corning) 35 3046)
- DMEM riche en glucose avec glutamine (Wisent 319-015-CL)
- FBS désactivé à la chaleur (Wisent 080 450)
- Pénicilline et streptomycine 100X (Wisent 450-201-EL)
- HEPES 1M (Wisent 330-050-EL)
- PBS (Wisent 311-010-CL)
- Trypsine avec .21 mM EDTA (Wisent 325-043-CL)
- Opti-MEM (Life Technologies (ThermoFisher) 31985-070)
- PEI (Polysciences Inc. 23966-1)
- Tris base (Wisent 600-125-IK)
- BSA (Santa Cruz Biotech Inc. sc-2323)
- NaCl (EMD-Millipore SX0420-5)

- Compteur de cellules Orflo Moxi-Z (MXZ001) avec cassettes de comptage (Orflo MXC012)
- Spectrophotomètre Tecan GENios Pro
- Apeline-13[Pyr¹] (Murza *et al*, 2012)
- Cellules HEK293
- Plasmides décrits dans la section « Génération des mutants de hAPJ »

Les cellules HEK293 sont toujours incubées à 37°C dans une atmosphère humidifiée contenant 5% de CO₂, sauf lorsque spécifié. Toutes les expériences utilisent uniquement des cellules HEK293.

Le milieu de culture appelé « standard » est du DMEM riche en glucose avec glutamine, supplémenté de 10% (v/v) de FBS désactivé à la chaleur, avec 1 UI/mL de pénicilline et 100 µg/mL de streptomycine.

Méthode de culture cellulaire de routine

Les cellules sont passées deux fois par semaine, en cycles alternés de 3 et 4 jours. Chaque passage, le milieu de culture est retiré, les cellules lavées doucement à la PBS, traitées pendant 2 min à la trypsine-EDTA, puis suspendues dans 10 mL de milieu standard. Dans le nombre requis de puits de culture de 10 cm préalablement remplis de 9 mL de milieu standard, 1 mL de la suspension cellulaire est mélangé et les puits sont de nouveau incubés jusqu'à leur utilisation.

Matériel spécifique à l'essai de liaison par compétition

- Apeline-13[Glp⁶⁵, Nle⁷⁵, Tyr⁷⁷][¹²⁵I] (Murza, Sainsily *et al*, 2017)
- EDTA (Sigma E5134)
- Plaques à 96 puits standard à fond conique
- Plaques à filtre à 96 puits (EMD Millipore MSFCN6B50)
- Compteur gamma PerkinElmer 2470 Wizard (80% d'efficacité)
- Tampon de collecte (10 mM Tris, 1mM EDTA, pH 7.4)
- Tampon de liaison (50 mM Tris, 0.2% BSA, pH 7.4)

Méthode des essais de liaison

Le milieu de culture des cellules est retiré, puis les cellules lavées doucement à la PBS, traitées pendant 2 min à la trypsine-EDTA, puis suspendues dans 8 mL de milieu de culture. Une petite partie de ce milieu est dilué 1/10 et la concentration de cellules est évaluée. Pour chaque mutant testé, un puits de culture de 10 cm estensemencé de 3.0×10^6 cellules et incubé 24 h.

Vers la fin de cette incubation, le réactif de transfection est préparé. 5 µg d'ADN plasmidique de 2HA-hAPJ et ses mutants sont mélangés à chacun 500 µL d'Opti-MEM préalablement réchauffé à 37°C. Après 5 min d'incubation à TP, 15 µL de PEI 1µg/µL est ajouté au mélange et agitée vigoureusement au vortex, suite à quoi le réactif de transfection est incubé 30 min à TP.

Le réactif de transfection est ensuite ajouté aux cellules incubées la veille et le milieu de culture est doucement agité pour bien le répandre, puis les cellules sont incubées de nouveau pendant 48 h.

Une fois l'incubation terminée, le milieu de culture est retiré, les cellules doucement rincées à la PBS, puis stockées à -80°C jusqu'à leur utilisation.

Le jour de l'expérience, des plaques à filtre sont remplies avec 75 µL par puits avec du PEI 1 µg/µL et gardées à 4°C jusqu'à utilisation. Chaque expérience considère huit dilutions 1/10 d'apeline-13 effectuées en triplicata pour chaque mutant, avec une place réservée au WT en guise de contrôle. Les cellules sont lysées par choc thermique de -80°C à 37°C pendant 1 à 2 min, puis rapidement mises sur glace. Toutes les étapes suivantes impliquant les cellules sont effectuées sur glace, sauf en cas de mention contraire.

Dans chaque puits de culture, 5 mL de tampon de collecte est ajouté, le fond du puits est gratté, la suspension de cellules lysées collectées dans un tube et le tout est centrifugé à 3500g pendant 15 min à 4°C. Le surnageant est retiré, puis le culot suspendu dans 1 mL de tampon de collecte.

Dans des plaques à 96 puits standard, des dilutions d'apeline-13[Pyr¹] sont préparées dans du tampon de liaison, de façon à couvrir des concentrations allant de 10⁻⁵ M à 10⁻¹² M dans le mélange final.

Le mélange du radioligand est préparé en mélangeant de l'apeline-13[Glp⁶⁵, Nle⁷⁵, Tyr⁷⁷][¹²⁵I] dans du tampon de liaison sur glace, de façon à obtenir un mélange à radioactivité de 875 cpm/μL.

Une fois les préparations d'apeline complétées, les préparations de membrane sont diluées dans 3 mL de tampon de liaison et agitées vigoureusement au vortex.

Dans chaque dilution d'apeline-13[Pyr¹], 120 μL de préparation de membrane et 40 μL de préparation de radioligand sont ajoutés, puis les mélanges sont incubés à TP pendant 1 h.

Suite à l'incubation, les plaques à filtre sont récupérées et le PEI aspiré. Les mélanges réactionnels sont ensuite pipetés dans les puits des plaques à filtre, le liquide aspiré, puis les filtres nettoyés trois fois au tampon de liaison préalablement mis sur glace. Les filtres sont ensuite collectés et la radioactivité analysée avec un compteur gamma.

Après trois expériences différentes pour chaque mutant, la moyenne de chaque mesure de chaque expérience est calculée et les moyennes des trois expériences sont analysées avec l'aide du logiciel GraphPad Prism en utilisant la moyenne ± SEM des trois expériences combinées effectuées en triplicata (donc la moyenne des observations et son écart type). La régression s'effectue en utilisant l'équation semi-logarithmique de liaison par compétition, avec E_{max} et E_{min} normalisés à 100% et 0%, respectivement :

$$E(\log[L]) = E_{max} + (E_{min} - E_{max}) \frac{10^{\log[L]}}{IC_{50} + 10^{\log[L]}}$$

À défaut d'obtenir un R² ajusté de plus de 0,95, il est présumé que le mutant satisfait une de ces deux conditions : il lie l'apeline radioactive à un K_D trop élevé pour que la liaison soit détectée avec fiabilité, ou encore il ne lie tout simplement pas l'apeline dans la plage des concentrations testées.

Matériel spécifique aux ELISAs de surface

- Poly-l-lysine 0,1% (m/w) (Sigma)

- Plaques de culture cellulaire à 24 puits (Corning 3526)
- Formaldéhyde 16% (m/v) (Thermo Fisher Scientific 28908)
- mAb de rat anti-HA-peroxydase (Roche Diagnostics 12 013 819 001 clone 3F10)
- TMB en solution (Sigma T4444)
- HCl 2N
- TBS (20 mM Tris, 150 mM NaCl, pH 8.0)
- Plaques 96 puits à fond transparent

Méthode des ELISAs de surface

La méthode est la même que pour les essais de liaison, jusqu'à l'ajout du réactif de transfection aux cellules. Suite à cet ajout, les cellules sont plutôt incubées 24h.

À la fin de la période d'incubation, le fond des plaques de culture à 24 puits est enduit de 300 μ L de poly-l-lysine. Le tout est incubé 10 minutes, puis la poly-l-lysine est récupérée et rangée pour une utilisation ultérieure et les plaques sont séchées pendant 30 minutes.

Pendant ce temps, le milieu de culture des cellules est retiré, les cellules sont doucement nettoyées à la PBS et traitées à la trypsine-EDTA pendant 2 min, puis suspendues dans du milieu standard.

Du milieu de culture standard, puis $3,5 \times 10^5$ cellules sont insérées dans chaque puits traité à la poly-l-lysine, à raison de 3 puits par mutant, avec un espace réservé pour le WT et le plasmide « Mock », puis le tout est incubé 24h.

Une fois l'incubation terminée, le milieu de culture est retiré et les cellules sont fixées avec de la TBS contenant 2,7% (m/v) de formaldéhyde pendant 5 min. Les cellules fixées sont lavées doucement deux fois à la TBS, suite à quoi la surface des puits est bloquée avec une solution de TBS et 1% (m/v) BSA pendant 30 min. La solution de blocage est ensuite retirée et les cellules exposées à du mAb de rat anti-HA dilué 1/1000 dans de la tBS et 1% (m/v) BSA pendant 3 h à TP.

Le mélange d'anticorps est retiré à la fin de la période d'incubation et les puits sont lavés doucement deux fois à la TBS. 250 μ L de solution de TMB sont ensuite ajoutés à chaque puits, le tout incubé 5 min, puis la réaction arrêtée avec 250 μ L de HCl 2N. 200 μ L de chaque réaction est pipeté dans une plaque 96 puits et l'absorbance à 450 nm est ensuite

lue. Chaque mutant est comparé au WT et au « Mock » dans au moins trois expériences séparées, avec le WT et le Mock normalisés respectivement à 100% et 0% d'expression, respectivement. Les mutants sont ensuite comparés au « Mock » à l'aide d'une ANOVA, afin d'identifier les mutants qui diffèrent du « Mock » en expression. Ces mutants sont considérés exprimés en surface et retenus pour les essais BRET.

Matériel spécifique aux essais BRET

- Plaques 96 puits blanches opaques pour culture cellulaire (Falcon (Corning) 353296)
- HBSS (Wisent 311-512-CL)
- Cœlenterazine 400A (Gold Biotechnology C-320-10)
- ADN de sperme de saumon

Méthode des essais BRET

Le milieu de culture des cellules est retiré, puis les cellules lavées doucement à la PBS, traitées pendant 2 min à la trypsine-EDTA, puis suspendues dans 8 mL de milieu de culture. Une petite partie de ce milieu est dilué 1/10 et la concentration de cellules est évaluée. Pour chaque mutant testé, un puits de culture de 10 cm contenant du milieu standard tamponné avec 20 mM d'HEPES estensemencé de 6.0×10^6 cellules et incubé 24 h. Pour les essais BRET, le milieu standard, peu importe où il est utilisé, est toujours tamponné avec 20 mM d'HEPES.

À la fin de la période d'incubation, des réactifs de transfection sont préparés, à raison de un par condition, dont le WT fait nécessairement partie, dans des tubes coniques de 5 mL et incubés 30 min à TP. Les proportions sont les suivantes :

- Pour les essais $G_{\alpha i1}$: 260 μ L d'Opti-MEM, 1245 ng d'ADN plasmidique de 2HA-hAPJ, 120 ng d'ADN plasmidique de $G_{\alpha i1}$ -RLucII, 750 ng d'ADN plasmidique de $G_{\beta 1}$, 750 ng d'ADN plasmidique de $G_{\gamma 1}$ -GFP₁₀, 135 ng d'ADN de sperme de saumon et 9 μ L de PEI à 1 μ g/ μ L.
- Pour les essais β -arrestine 2 : 260 μ L d'Opti-MEM, 750 ng d'ADN plasmidique de 2HA-hAPJ, 120 ng ADN plasmidique de β -arrestine 2-RLucII, 750 ng d'ADN

plasmidique de CAAX-rGFP, 1380 ng d'ADN de sperme de saumon et 9 µL de PEI à 1 µg/µL.

Pendant l'incubation du réactif, les cellules sont lavées doucement à la PBS, traitées 2 min à la trypsine-EDTA, puis l'ensemble des cellules suspendu dans du milieu de culture. Les cellules sont comptées, puis diluées de façon à obtenir un milieu de culture avec $3,7 \cdot 10^5$ c/mL. 2,7 mL de cette suspension sont ajoutés à chaque préparation de réactif de transfection et le tout est mélangé doucement pour répartir le réactif dans toute la suspension. Chacun de ces mélanges est ensuite pipeté, à raison de 100 µL par puits, dans des plaques 96 puits blancs opaques pour culture cellulaire et le tout est incubé pendant 48 h. Les incubations subséquentes sont toutes effectuées à TP dans une atmosphère normale. Une fois l'incubation terminée, le milieu de culture est retiré, remplacé avec 80 µL de HBSS tamponné avec 20 mM d'HEPES et le tout incubé 1 h. Pendant ce temps, les dilutions d'apeline-13[Pyr¹] sont préparées dans le même milieu, en changeant d'embouts à chaque dilution, pour obtenir des concentrations finales allant de 10^{-5} M à 10^{-11} M, plus un blanc sans apeline.

Une fois l'incubation terminée, 10 µL de Coelenterazine 400A à 50 mM est ajoutée dans chaque puits et le tout incubé 10 min. Les dilutions d'apeline-13[Pyr¹] sont ensuite ajoutées et le tout incubé encore 10 min. La bioluminescence est ensuite lue à 525 nm pour la GFP10 et rGFP et sous 450 nm pour la RLucII avec un spectrophotomètre. Le rapport BRET (comptes d'émissions GFP/RLucII) est ensuite calculé et les paramètres d'une courbe dose-réponse non-coopérative semi-logarithmique sont optimisés pour chaque condition, avec un minimum de trois expériences indépendantes par condition, avec les paramètres du WT utilisés pour normaliser chaque expérience :

$$E(\log[L]) = E_{min} + (E_{max} - E_{min}) \frac{10^{\log[L] + pEC_{50}}}{1 + 10^{\log[L] + pEC_{50}}}$$

La moyenne des observations et son écart type (SEM) sont utilisés pour effectuer la régression. Afin d'identifier avec certitude les mutants présentant une activité constitutive, une ANOVA est effectuée sur les valeurs d'activation normalisée des mutants sans apeline-

13, suivie du test de Dunnett pour comparaisons multiples. La même chose est effectuée sur les valeurs d'activation normalisée à 10^{-5} M d'apeline-13 pour identifier les mutants dont l'activité maximale est supérieure ou inférieure à celle du WT.

Résultats

Analyse des systèmes hAPJ-apeline et hAPJ-apo

Analyse théorique du système hAPJ-apeline

Il est observé que le système hAPJ-apeline optimal est constitué de deux composantes indépendantes et de quatre ensembles conformationnels distincts.

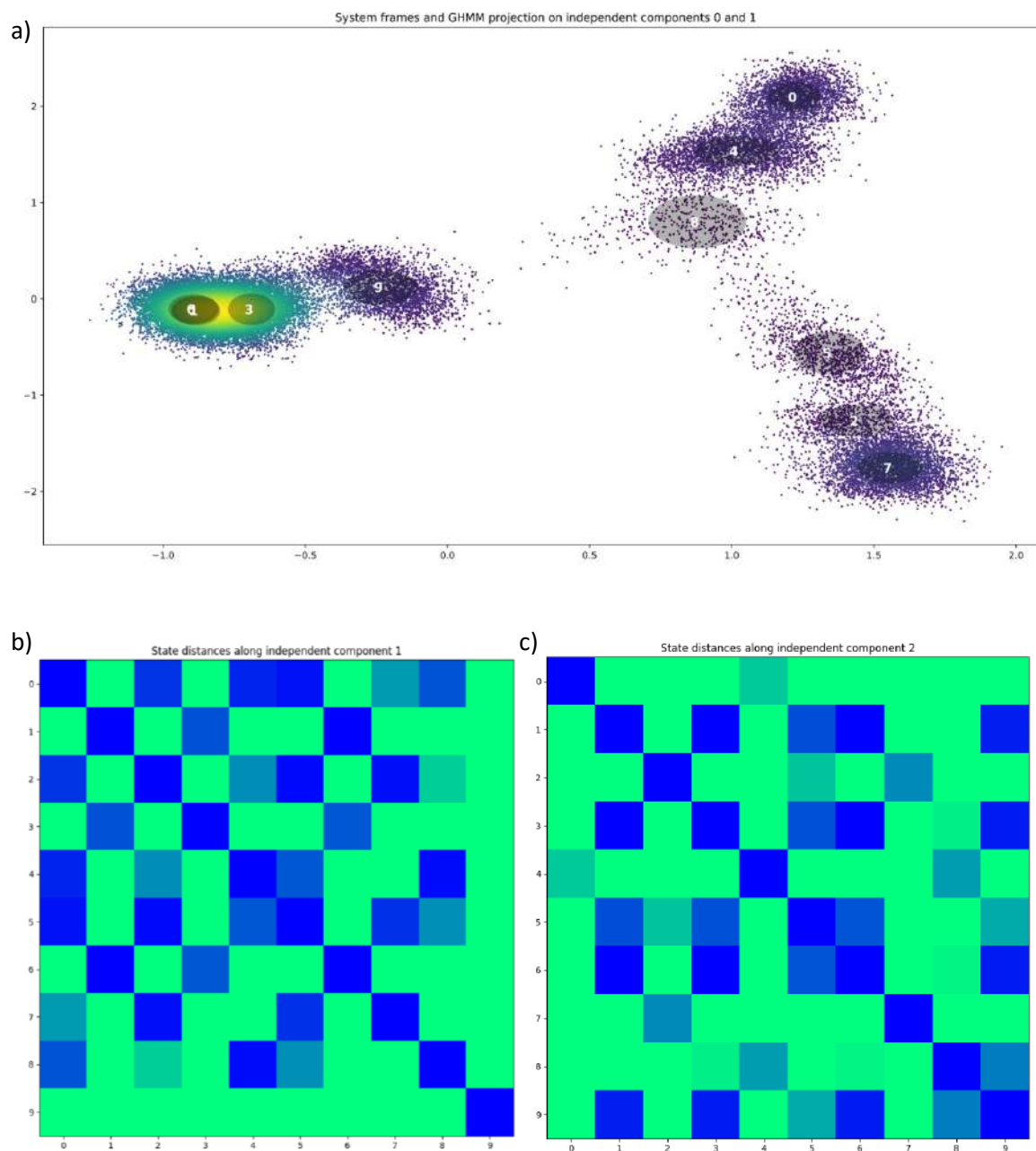


Figure 15: Résultats du programme d'analyse de base du système hAPJ-WT-apeline

a) Illustration de la distribution des conformations explorées par l'ensemble des trajectoires du système hAPJ-apeline. Deux composantes indépendantes et dix états (numérotés de 0 à 9) sont nécessaires pour bien décrire le système. b) Représentation en *heat map* de la distance de Bhattacharyya entre les différents états le long de la composante indépendante 1. Les couleurs vont de bleu (0) à vert (plus de 1,982, ou erreur de classification inférieure à 2,5%). On remarque que l'état 9 est complètement séparé du reste du système via cette composante. On remarque également que, hormis l'état 9, le système est réparti en deux ensembles conformationnels par cette composante : les états 0, 2, 4, 5, 7 et 8 forment un ensemble et les états 1, 3 et 6 en forment un autre. Ces séparations correspondent à ce qui est observé en a). c) Représentation en *heat map* de la distance de Bhattacharyya entre les différents états le long de la composante indépendante 2. On remarque ici que l'ensemble des états 0, 2, 4, 5, 7 et 8 défini par la composante indépendante 1 se sépare de nouveau : les nouveaux sous-ensembles deviennent 2, 5 et 7, ainsi que 0, 4 et 8. Ces séparations correspondent également à ce qui est observé en a).

Les deux composantes indépendantes nécessaires à la description de la distribution conformationnelle du système à l'équilibre présentent une signature qui met en évidence le rôle des différentes régions de la protéine dans les changements conformationnels observés par l'analyse des distances de Bhattacharyya. Dans les deux cas, les signatures les plus intenses se concentrent au niveau des boucles extracellulaires et intracellulaires, ainsi qu'autour du Hydrophobic Hindering Mechanism et du Major H-Bond Network. Ces signatures, combinées à l'observation des conformations des différents états, témoignent de deux changements dans ces régions : au niveau du MHBN, un réarrangement important se manifeste par une dénaturation locale du TM7 et au niveau du HHM, le résidu F257 change de conformation, passant d'une orientation qui le rapproche du TM7 à une qui le rapproche du TM5.

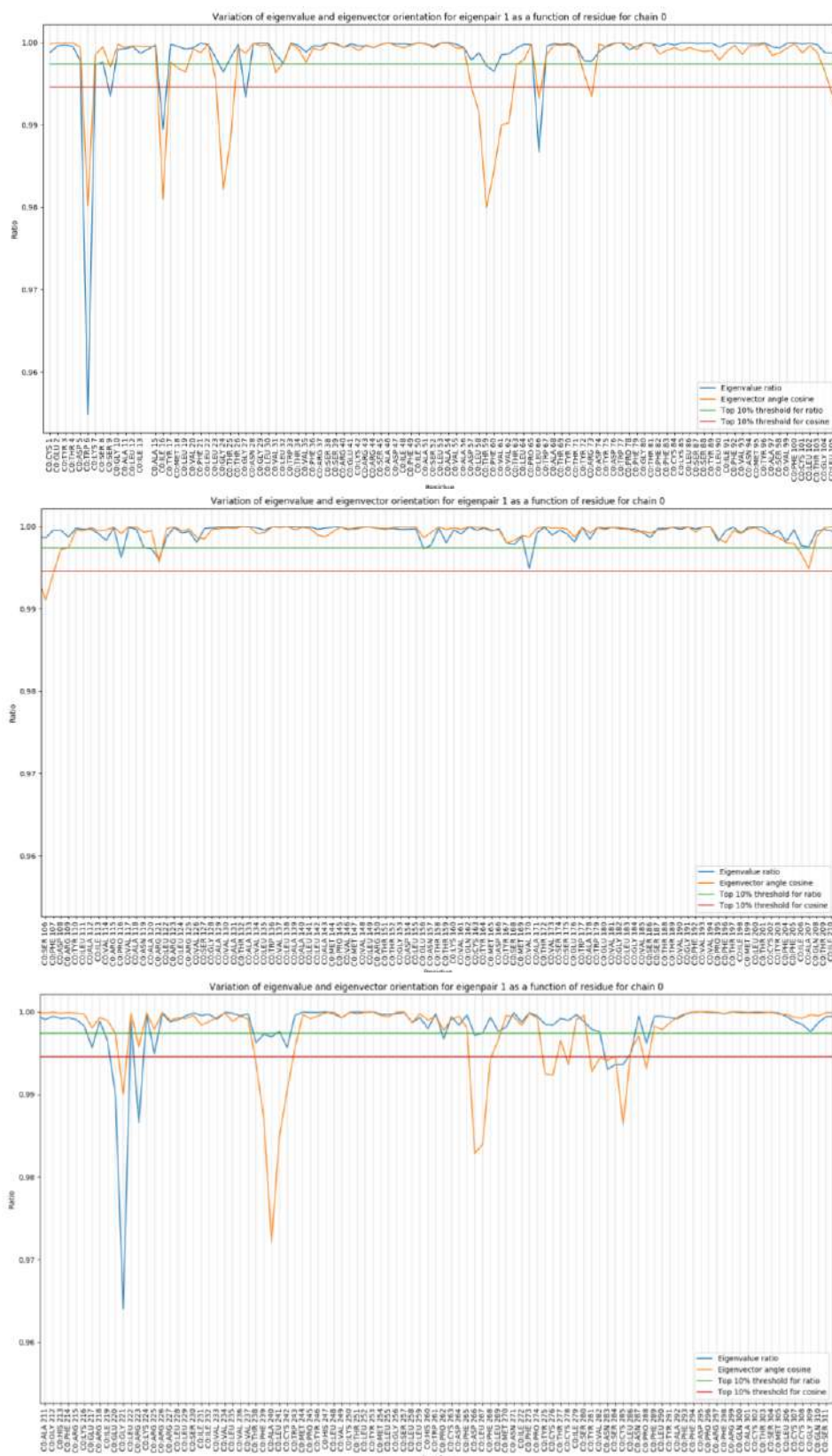


Figure 16 : Résultat de l'analyse par sous-matrices principales de la composante indépendante 1 du système hAPJ-WT-apeline. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée. Le résidu P32 est absent du graphique car, en raison d'un bug de MDTraj, il est identifié comme P340. Ce résidu n'influe pas de façon significative sur la composante indépendante 1.

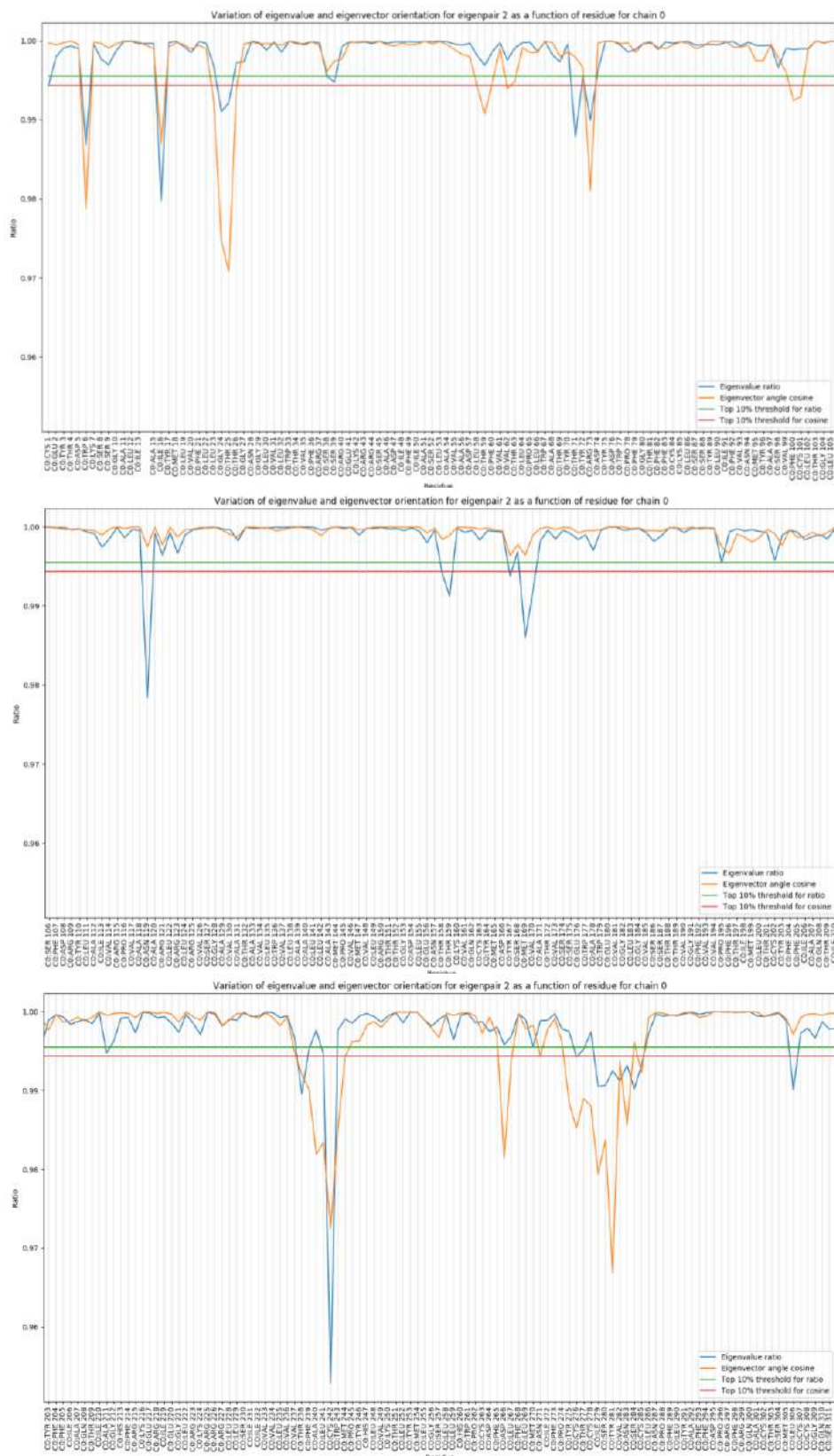


Figure 17 : Résultat de l'analyse par sous-matrices principales de la composante indépendante 2 du système hAPJ-WT-apeline

Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée. Le résidu P32 est absent du graphique car, en raison d'un bug de MDTraj, il est identifié comme P340. Ce résidu n'influe pas de façon significative sur la composante indépendante 2.

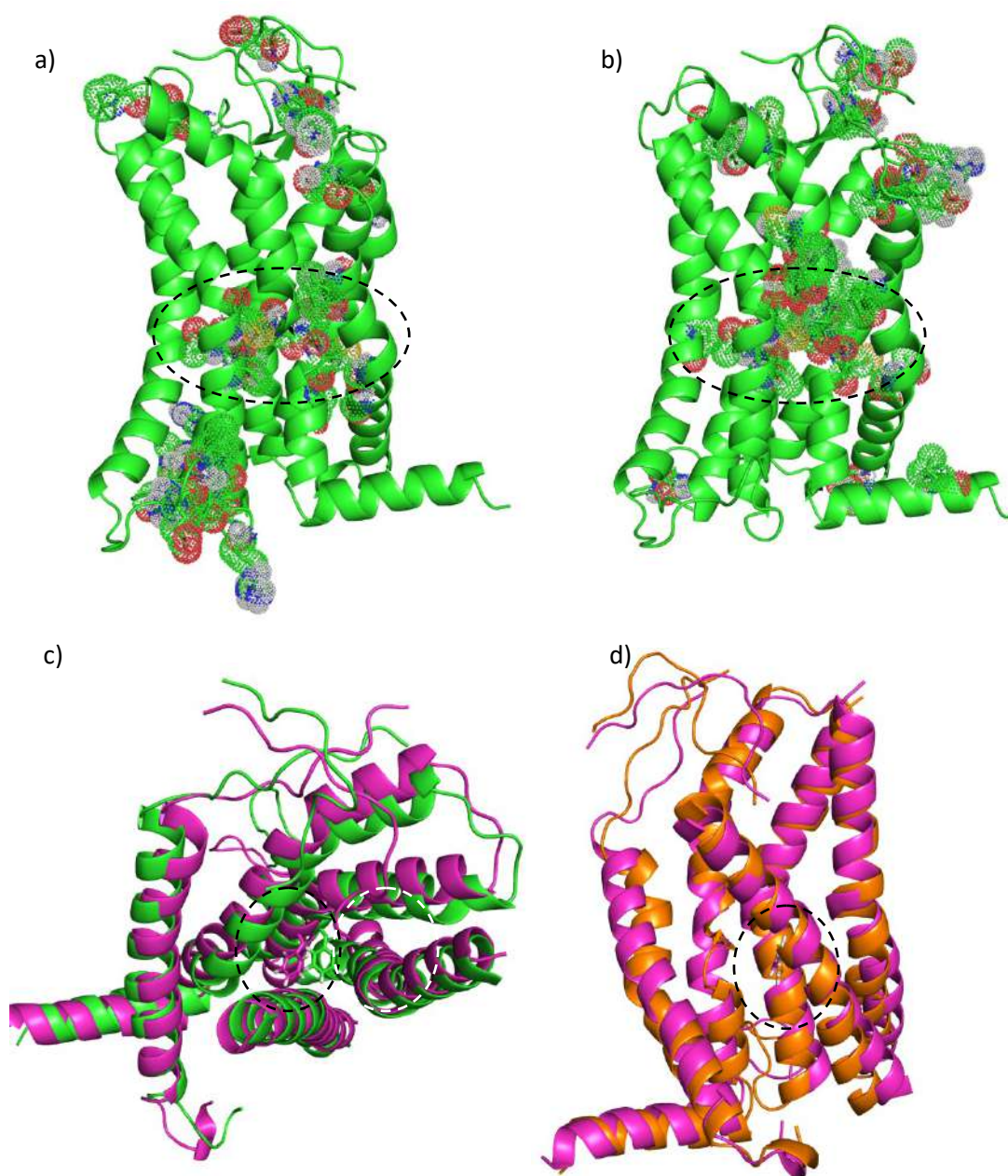
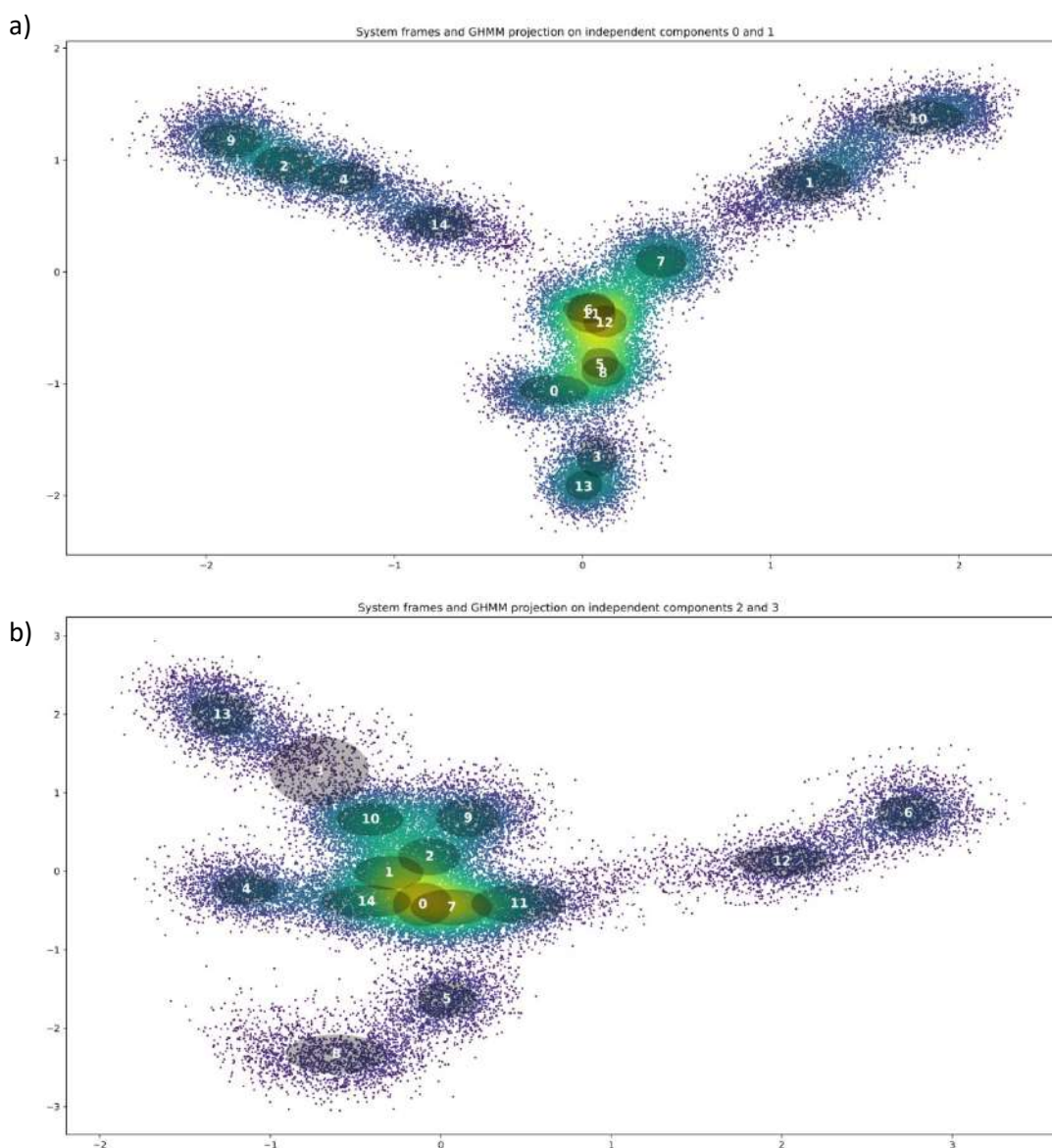


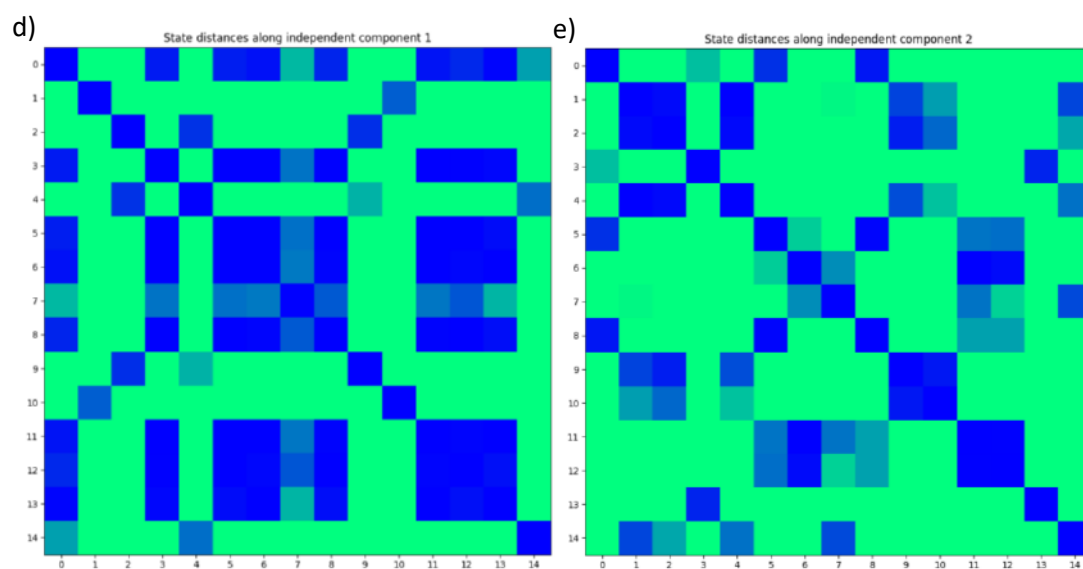
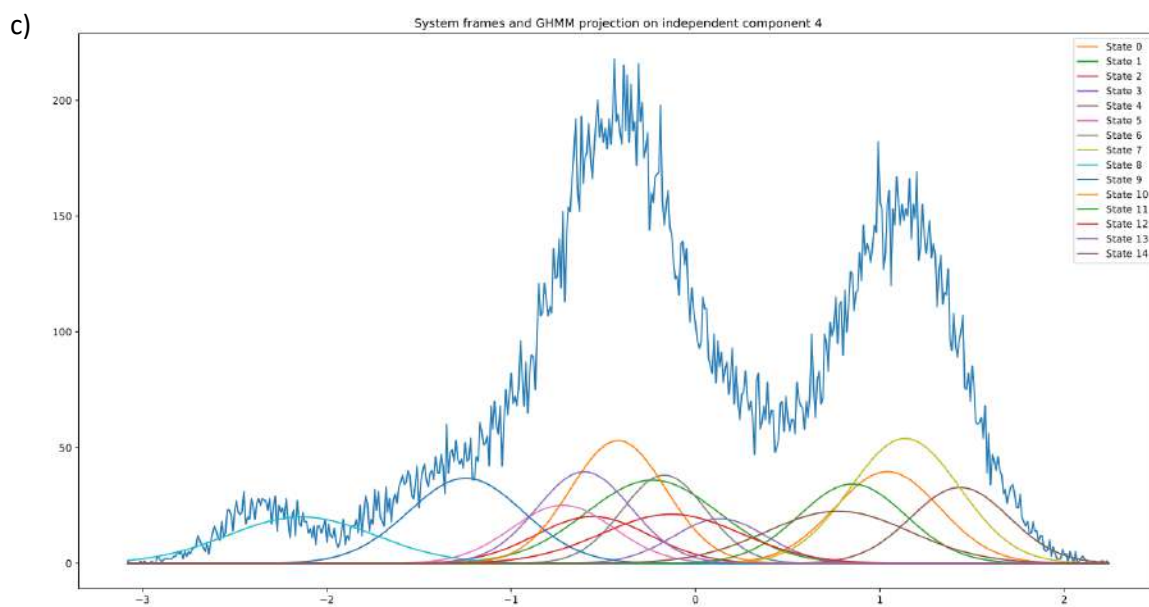
Figure 18 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-WT-apeline

a) Répartition sur hAPJ des résidus critiques de la composante indépendante 1 indiqués par l'analyse par sous-matrices principales du système hAPJ-apeline. La section encadrée est centrée autour du MHBN et inclut F257 du HHM. Les autres résidus identifiés font partie des boucles intracellulaires et extracellulaires. b) Répartition sur hAPJ des résidus critiques de la composante indépendante 2 indiqués par l'analyse par sous-matrices principales du système hAPJ-apeline. La section encadrée est aussi centrée autour du MHBN et inclut F257 et P213^{5,50}. Les autres résidus identifiés font partie des boucles intracellulaires et extracellulaires. c) Représentation graphique du changement conformationnel reliant les états 1, 3 et 6 (magenta) aux états 0, 4 et 8 (vert). Une distorsion importante du TM7 peut être observée (encadré en noir), ainsi qu'un changement conformationnel de F257 (encadré en blanc). d) Représentation graphique du changement conformationnel reliant les états 1, 3 et 6 (magenta) aux états 2, 5 et 7 (orange). Une dénaturation locale de la structure hélicale du TM7 peut être observée (encadré en noir). F257 ne change pas de conformation dans cette transition.

Analyse théorique du système hAPJ-apo

Il est observé que le système hAPJ-apo optimal est constitué de cinq composantes indépendantes et de plusieurs ensembles conformationnels distincts. Cependant, leur séparation est moins claire que pour le système hAPJ-apeline. Les graphiques d'analyse par sous-matrices principales du système se trouvent à l'Annexe IV. De par sa proximité de la moyenne des conformations du système sans apeline, l'état 11 est utilisé comme référence pour comparer les états de tous les systèmes.





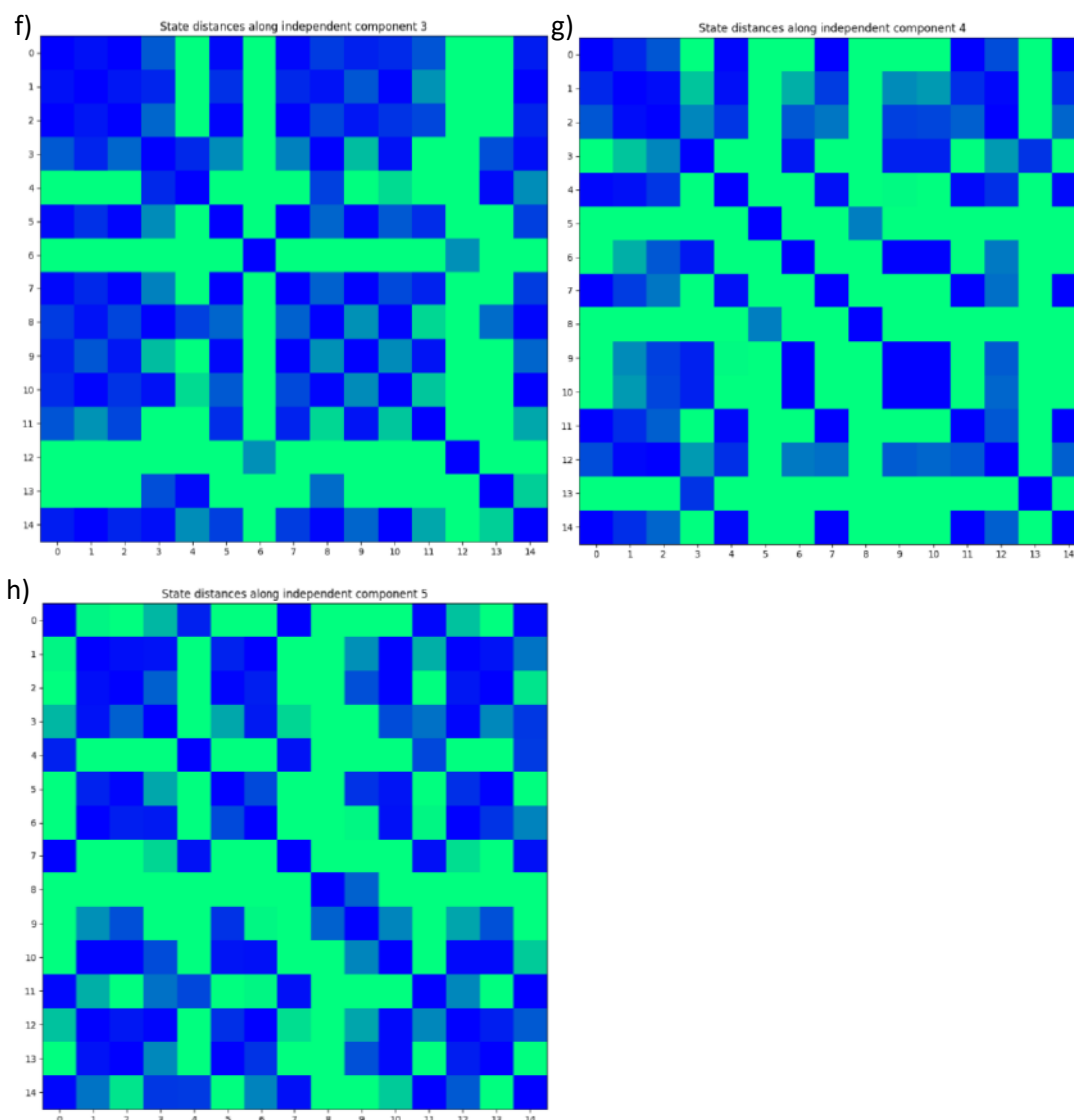
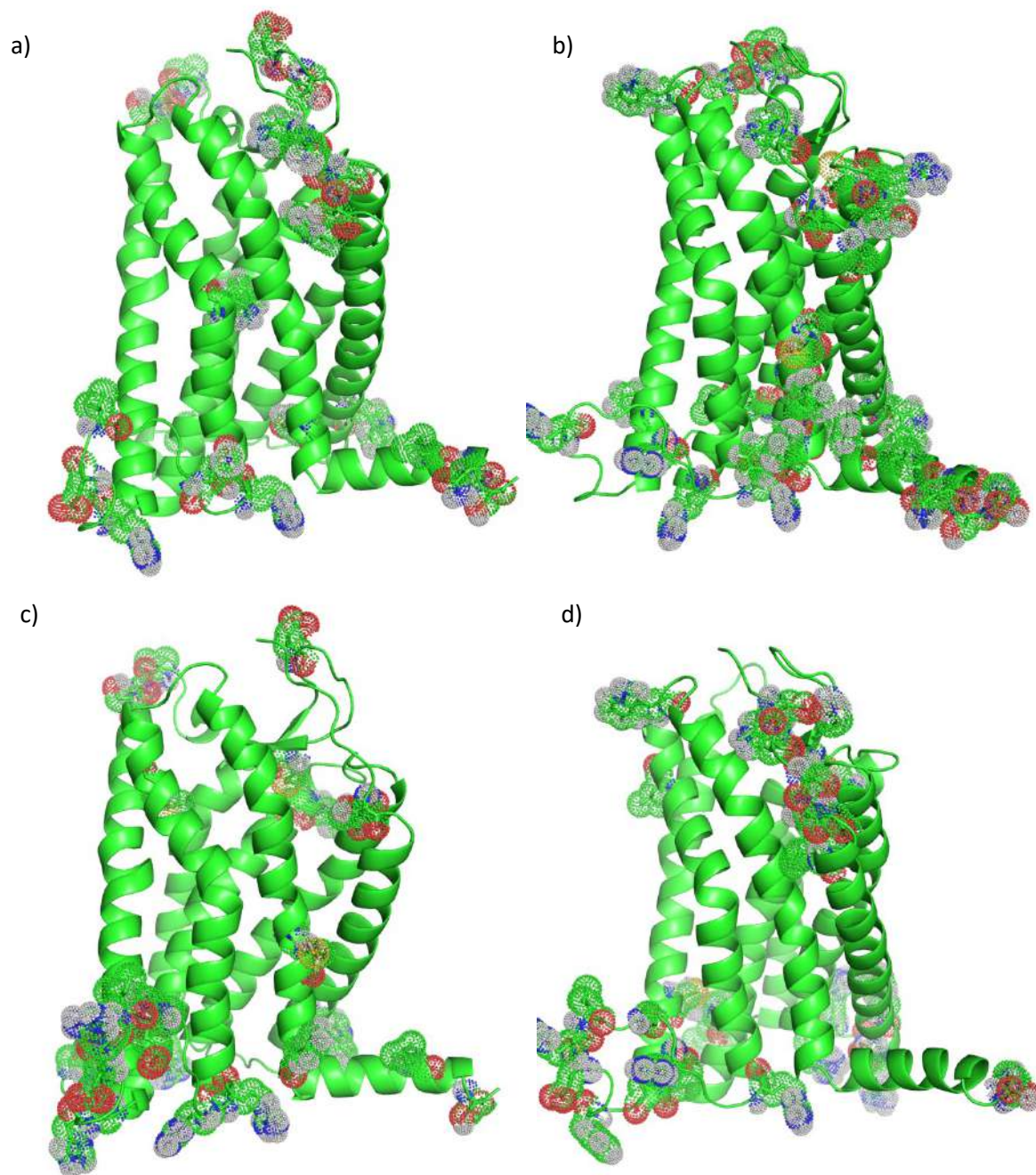


Figure 19 : Résultats du programme d'analyse de base du système hAPJ-WT-apo

a, b, c) Illustration de la distribution des conformations explorées par l'ensemble des trajectoires du système hAPJ-apeline. Cinq composantes indépendantes et quinze états (numérotés de 0 à 14) sont nécessaires pour bien décrire le système. d, e, f, g, h) Représentation en *heat map* de la distance de Bhattacharyya entre les différents états le long des différentes composantes indépendantes. Les couleurs vont de bleu (0) à vert (plus de 1,982, ou erreur de classification inférieure à 2,5%). Les différents ensembles conformationnels dont la distinction du reste du système ne requiert qu'une composante sont les ensembles contenant les états 1 et 10 (composante 1), 6 et 12 (composante 3) et 5 et 8 (composante 4). Les autres états sont définis par un cumul des différents changements conformationnels représentés par les composantes indépendantes du système. Par exemple, l'ensemble contenant les états 2, 4, 9 et 14 est séparé des autres grâce aux composantes 1 (états 1, 7 et 10 vs 2, 4, 9 et 14) et 2 (état 0 vs 14).



e)

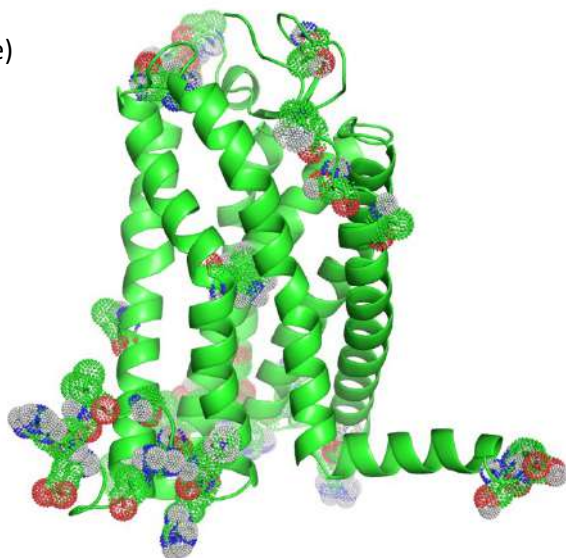


Figure 20 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-WT-apo a-e) Répartition sur hAPJ des résidus critiques des composantes indépendantes 1 à 5 indiqués par l'analyse par sous-matrices principales du système hAPJ-apo. Le motif NPxxY semble impliqué dans le changement conformationnel associé à la composante 2 et tryptophane du motif WxP semble impliqué dans ceux des composantes 1 et 5. Hormis ces deux régions, les autres résidus critiques font partie des boucles intracellulaires et extracellulaires.

Design de mutants de hAPJ

Considérant les résultats théoriques disponibles sur les systèmes hAPJ-apeline et hAPJ-apo, la région à cibler est celle entourant ou constituant le MHBN et le HHM. Le design de mutants est basé sur les résultats existants de la littérature, avec pour critère simple que le mutant doit altérer de façon importante la fonctionnalité du récepteur étudié. Une explication plus en détail des mutants identifiés se trouve à l'Annexe III. Un autre critère important est que, comme l'étude porte sur la dynamique du récepteur avec ou sans l'apeline-13, le site de la mutation ne doit pas se trouver dans le site de liaison de l'apeline-13. Ainsi, l'impact de la mutation sur la structure du site de liaison se limite, autant que possible, à l'impact qu'a la mutation sur l'équilibre des conformations du récepteur.

Analyse fonctionnelle des mutants sélectionnés

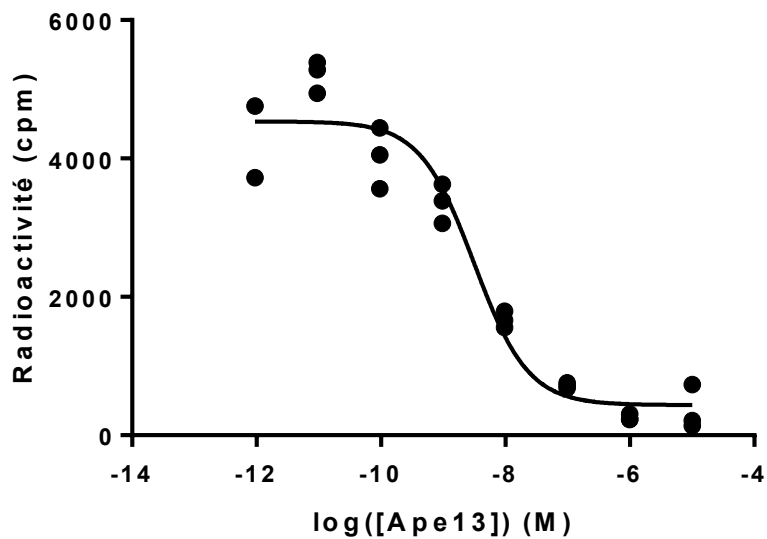
Essais de liaison par compétition

Il est observé que la majorité des mutants testés lient l'apeline-13 de façon observable selon un modèle de liaison non-coopératif.

Tableau 4 : Paramètres d'intérêt pour l'étude du déplacement de l'apeline radiomarquée par l'apeline non-radiomarquée par régression non-coopérative. * : Corrélation non-significative au sens de Pearson (relation Michaélienne), mais significative au sens de Spearman (relation monotone décroissante) ** : Corrélation non-significative au sens de Spearman N/A : Non-applicable (pas de régression Michaélienne significative)					
Mutant	$\log IC_{50} \pm \text{SEM}$	R^2 ajusté	Mutant	$\log IC_{50} (\pm \text{SEM})$	R^2 ajusté
WT	$-8,62 \pm 0,09$	0,97	L253Q	$-8,72 \pm 0,08$	0,98
D75N	N/A	0,01**	L253I	$-8,54 \pm 0,09$	0,97
N112G	$-8,51 \pm 0,12$	0,95	V254L	$-8,72 \pm 0,10$	0,96
N112F	N/A	0,82*	F257T	$-8,29 \pm 0,08$	0,97
N112W	N/A	0,31**	F257L	$-8,69 \pm 0,08$	0,98
V117L	$-8,82 \pm 0,08$	0,97	F257G	N/A	0,93*
L120R	N/A	0,94*	F257W	N/A	0,86*
L120K	$-8,68 \pm 0,07$	0,98	W261R	N/A	0,12**
L120A	$-9,17 \pm 0,10$	0,96	W261F	N/A	0,94*
D126N	$-8,82 \pm 0,10$	0,96	N301G	$-8,65 \pm 0,08$	0,98
D126V	$-8,65 \pm 0,08$	0,98	S302N	N/A	0,76*
L253F	$-8,66 \pm 0,07$	0,98	Y309F	N/A	0,92*

Selon le mutant étudié, les courbes de déplacement présentaient une plus ou moins grande variabilité intra-expérience. Par exemple, voici une courbe de déplacement représentative du WT et une autre du mutant W261F, dont la haute variabilité intra-expérience est responsable de la mauvaise régression finale :

a) **Données représentatives d'une expérience isolée de liaison par compétition du WT**



b) **Données d'une expérience isolée de liaison par compétition du mutant W 261 F**

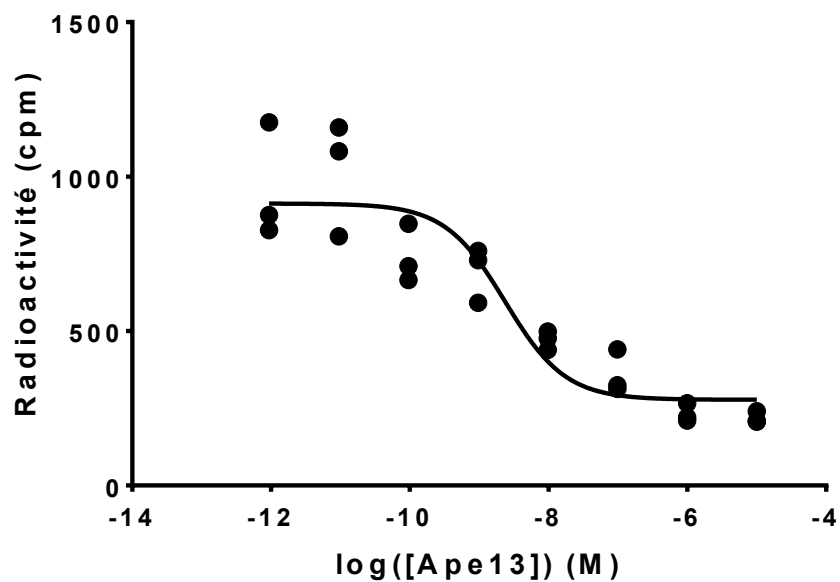


Figure 21 : Données expérimentales représentatives d'un essai de liaison par compétition Sur a) hAPJ-WT et b) hAPJ-W261F. L'erreur de la mesure est beaucoup plus prononcée dans les essais chez W261F en raison d'un signal plus faible, ce qui explique la mauvaise régression.

Les figures de la régression finale pour chaque mutant se trouvent à l'Annexe I.

ELISAs de surface

Il est observé que la totalité des récepteurs est exprimée à la surface de la cellule, avec un niveau de risque inférieur à 5%. Cependant, les mutants L120R et L120A ne sont pas exprimés à la surface avec un niveau de risque inférieur à 1%, ce qui indique que leur taux d'expression en surface est faible. La p-value est calculée à l'aide du test de comparaisons multiples de Dunnett suivant une ANOVA avec le logiciel GraphPad Prism.

Il est également observé que la plupart des mutants semble avoir un niveau d'expression situé entre le « Mock » et le WT. Pour les besoins du travail, un niveau d'expression normalisé de plus de 50% est considéré similaire à celui du WT.

Tableau 5 : Paramètres d'intérêt pour l'étude de l'expression des mutants à la surface de la cellule par ELISA de surface.

La p-value représente la probabilité que le niveau d'expression du mutant diffère de celui du « Mock »

* : Niveau d'expression égal à 0% avec $\alpha = 0.01$

Mutant	Expression normalisée (% \pm SEM)	p-value	Mutant	Expression normalisée (% \pm SEM)	p-value
WT	100	0	L253Q	93 \pm 9	0,0001
Mock	0	1	L253I	86 \pm 6	0,0001
D75N	111 \pm 2	0,0001	V254L	67 \pm 6	0,0001
N112G	83 \pm 12	0,0001	F257T	69 \pm 2	0,0001
N112F	39 \pm 1	0,0001	F257L	41 \pm 3	0,0001
N112W	72 \pm 7	0,0001	F257G	74 \pm 10	0,0001
V117L	49 \pm 6	0,0001	F257W	116 \pm 4	0,0001
L120R	22 \pm 3	0,0397*	W261R	123 \pm 5	0,0001
L120K	36 \pm 3	0,0001	W261F	51 \pm 9	0,0001
L120A	26 \pm 2	0,0162*	N301G	59 \pm 11	0,0001
D126N	93 \pm 3	0,0001	S302N	46 \pm 1	0,0001
D126V	61 \pm 1	0,0001	Y309F	82 \pm 3	0,0001
L253F	70 \pm 8	0,0001			

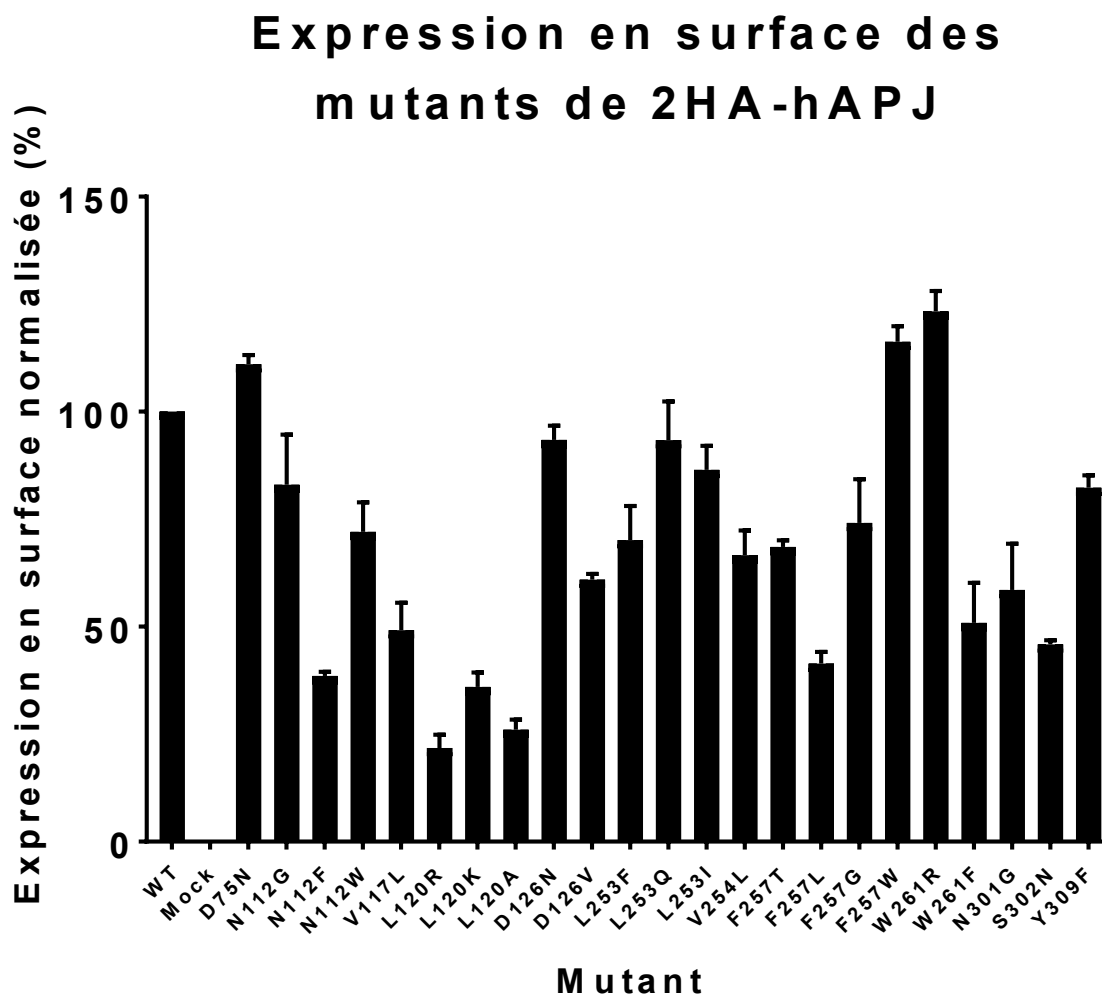


Figure 22 : Expression relative des mutants de 2HA-hAPJ à la surface de la cellule mesurée par ELISA de surface. Sans être quantitatif, cet essai permet de mesurer le niveau d'expression relatif en surface de chaque mutant. Les mutants L120R et L120A sont ceux avec le niveau d'expression en surface le plus faible. La plupart des mutants ont un niveau d'expression en surface supérieur ou égal à 50% de 2HA-hAPJ-WT. Seuls trois mutants ont un niveau d'expression en surface légèrement plus élevé que 2HA-hAPJ-WT : 2HA-hAPJ-D75N, 2HA-hAPJ-F257W et 2HA-hAPJ-W261R.

Essais BRET

Suite au test de Dunnett sur l'activité basale normalisée des mutants de 2HA-hAPJ, il est observé que trois mutants présentent une activité constitutive pour le recrutement de la protéine $G\alpha_{i1}$: N112G, D126V et V254L. De même, il est observé que quatre mutants présentent une activité constitutive pour le recrutement de la β -arrestine 2 : N112G, N112W, L120K et L253Q. Il est important de remarquer que les résultats de BRET pour $G\alpha_{i1}$ présentent plus de bruit que les résultats de BRET pour la β -arrestine 2. En guise de comparaison, une expérience représentative de $G\alpha_{i1}$ et une autre de β -arrestine 2 sont montrées après les tableaux. Cette différence est due au faible écart de rapport BRET $G\alpha_{i1}$ (de l'ordre de 0,15) entre la forme stimulée et non-stimulée du récepteur de type sauvage, comparé à celui du BRET β -arrestine 2 (de l'ordre de 1,0). En raison des grandes incertitudes de mesure du BRET $G\alpha_{i1}$, il est hasardeux de tirer des conclusions sur l'exclusivité des mutants N112W, L120K et L253Q pour la β -arrestine 2 en ce qui concerne leur activité constitutive. En effet, un expérimentateur indépendant (Xavier Sainsily, labo Sarret, Université de Sherbrooke) a confirmé que le mutant L253Q, le plus prometteur, présente une activité constitutive pour $G\alpha_{i1}$. Cependant, du côté des mutants N112G, D126V et V254L, il est plus clair que N112G présente une activité constitutive autant pour $G\alpha_{i1}$ et pour la β -arrestine 2, tandis que D126V et V254L présentent une activité constitutive uniquement pour $G\alpha_{i1}$.

Tableau 6 : Paramètres d'intérêt pour l'étude du recrutement de la protéine $G\alpha_{i1}$ par les mutants de 2HA-hAPJ

* : Régression non-significative

† : Activité maximale différente de celle du WT (niveau de risque de 5%)

‡ : Activité basale différente de celle du WT (niveau de risque de 5%)

** : Activité constitutive d'intensité incertaine. Considéré par prudence comme non-constitutivement actif.

Mutant	$\log EC_{50}$	E_{min} (%)	E_{max} (%)	R^2 ajusté	Mutant	$\log EC_{50}$	E_{min} (%)	E_{max} (%)	R^2 ajusté
WT	-9,46	0	100	0,97	L253Q	-9,14	13,7**	68,2†	0,76*
D75N	-7,17	-21,3	65,0†	0,91	L253I	-9,26	26,1	107	0,92
N112G	-9,31	92,4‡	108	0,48*	V254L	-9,38	38,1‡	110	0,92
N112F	-9,20	18,8	80,3	0,97	F257T	-9,52	8,92	98,2	0,91
N112W	-8,57	31,9**	84,0	0,76*	F257L	-9,16	18,3	84,7	0,75*
V117L	-9,22	-9,48	87,6	0,95	F257G	-8,88	0,11	82,5	0,82*
L120R	-9,37	-5,96	28,6†	0,47*	F257W	-8,15	-9,28	73,3†	0,78*
L120K	-9,13	20,5**	79,5†	0,85*	W261R	-7,58	-21,3	74,5†	0,77*
L120A	-9,30	26,2	76,2†	0,81*	W261F	-9,34	10,8	99,6	0,93
D126N	-9,32	24,4	108	0,90	N301G	-9,43	26,5	103	0,89*
D126V	-9,60	59‡	117†	0,94	S302N	-9,24	2,17	87,0	0,91
L253F	-9,37	17,6	99,2	0,92	Y309F	-9,27	-6,99	97,1	0,92

Tableau 7 : Paramètres d'intérêt pour l'étude du recrutement de la protéine β -arrestine 2 par les mutants de 2HA-hAPJ

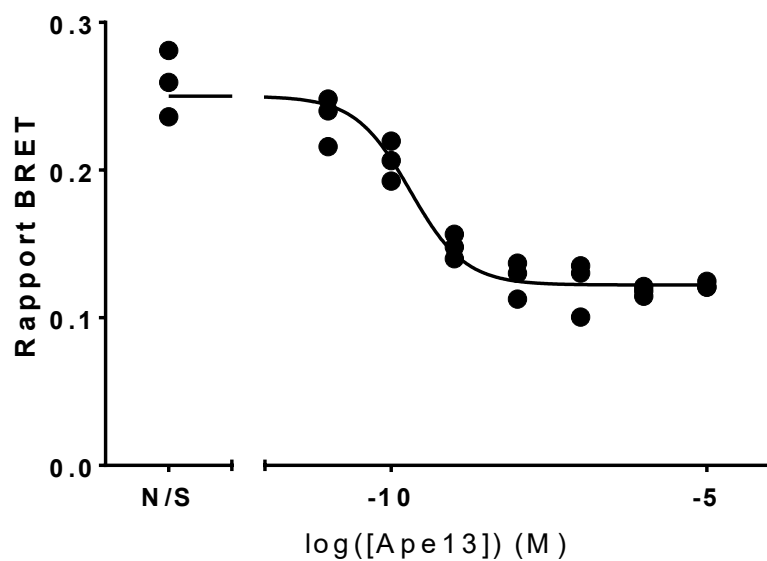
* : Régression non-significative

† : Activité maximale estimée différente de celle du WT (niveau de risque de 5%)

‡ : Activité basale différente de celle du WT (niveau de risque de 5%)

Mutant	$\log EC_{50}$	E_{min} (%)	E_{max} (%)	R^2 ajusté	Mutant	$\log EC_{50}$	E_{min} (%)	E_{max} (%)	R^2 ajusté
WT	-7,94	0	100	0,99	L253Q	-8,39	33,8‡	81,8†	0,86*
D75N	-6,14	0,36	45,8†	0,98	L253I	-8,09	-1,40	89,4	0,99
N112G	-7,91	85,7‡	97,7	0,21*	V254L	-8,40	-2,09	48,4†	0,99
N112F	-7,89	-3,35	18,1†	0,96	F257T	-8,07	-2,94	44,1†	0,97
N112W	-6,43	30,6‡	56,0†	0,92	F257L	-7,86	-4,49	23,0†	0,95
V117L	-7,88	-4,39	43,1†	0,98	F257G	-7,27	-3,92	19,3†	0,97
L120R	-8,89	3,60	5,17†	0,36*	F257W	-6,75	-0,54	10,4†	0,84*
L120K	-8,19	7,67‡	19,3†	0,92	W261R	-6,29	3,46	37,9†	0,98
L120A	-8,45	-0,97	9,31†	0,93	W261F	-8,10	-4,40	46,6†	0,98
D126N	-8,31	-1,49	75,4†	0,98	N301G	-8,65	-1,74	40,1†	0,98
D126V	-8,76	3,66	41,0†	0,96	S302N	-7,91	-4,76	29,4†	0,98
L253F	-8,40	-2,83	62,8†	0,98	Y309F	-8,09	-2,27	73,8†	0,90*

a) **Données brutes de rapport BRET**
représentatif d'un essai G_{i1}
sur 2HA-hAPJ WT



b) **Données brutes de rapport BRET**
représentatif d'un essai β -arrestine₂
sur 2HA-hAPJ WT

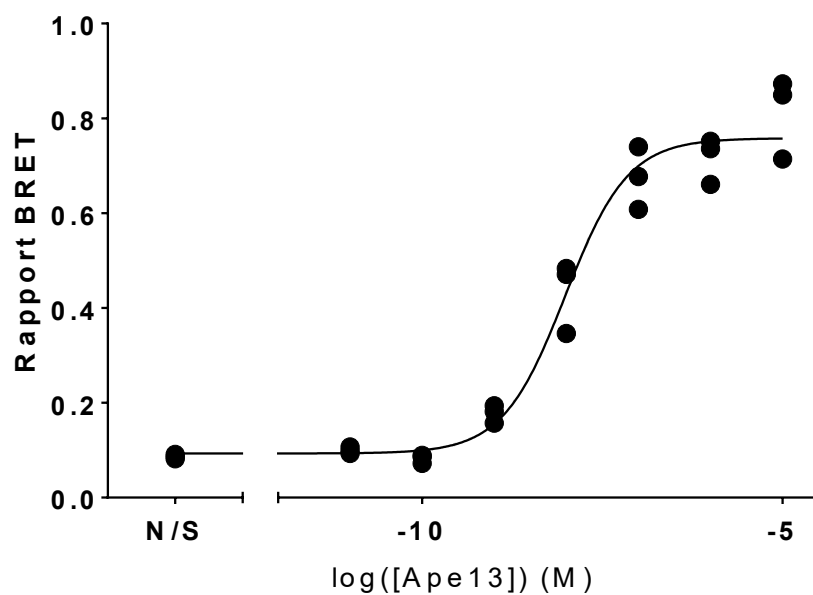


Figure 23 : Données brutes représentatives d'un essai BRET

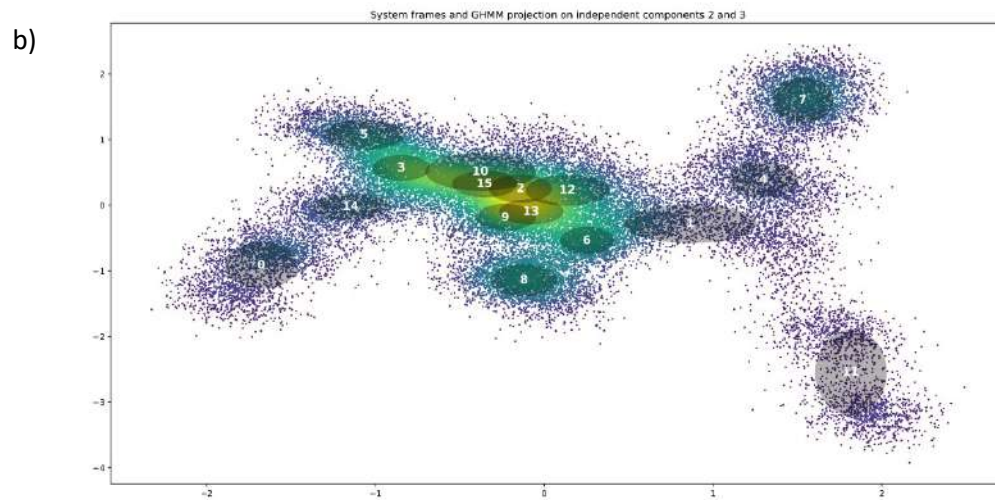
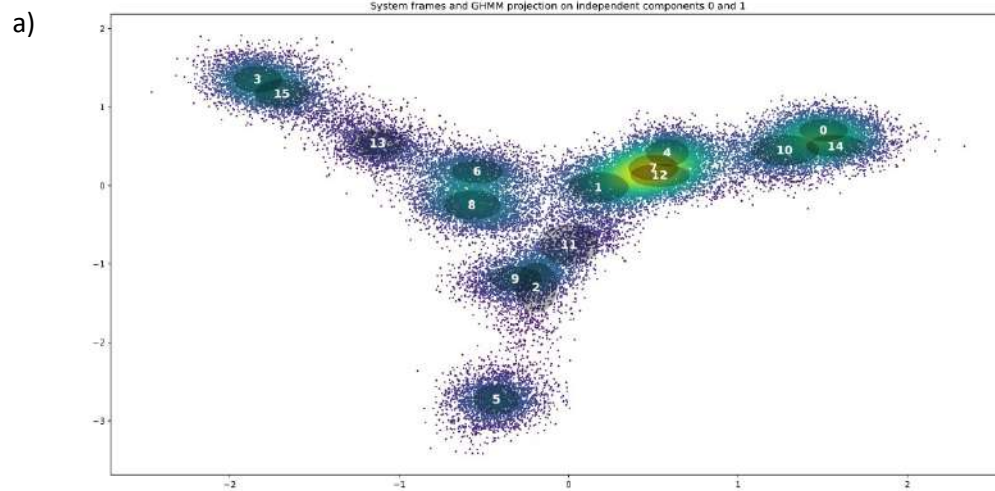
a) $G\alpha_{i1}$ et b) β -arrestine 2 sur 2HA-hAPJ-WT. Non seulement la variation du rapport BRET est plus faible pour l'essai $G\alpha_{i1}$, mais l'incertitude sur l'activité basale est plus grande. Ces deux observations mettent en garde contre une interprétation hâtive des résultats et justifient une contre-vérification par un expérimentateur indépendant.

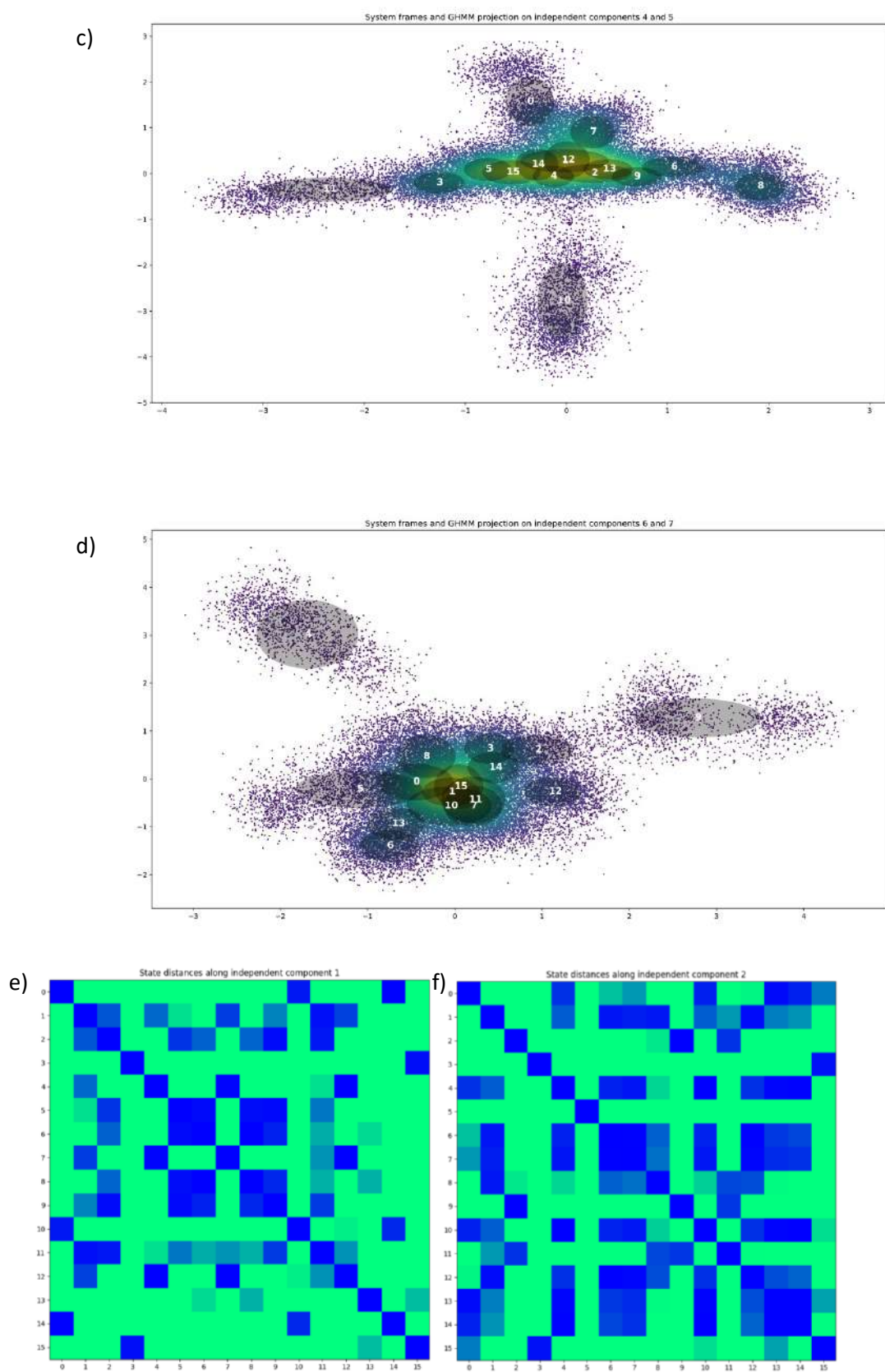
La totalité des données finales de BRET est présentée sous forme de graphiques dans l'Annexe II. Il est également à noter que pour certains mutants, l'observation de l'activité à 10^{-11} M est anormalement basse par rapport au reste des observations. La cause de cette anomalie est inconnue, mais comme l'observation sans apeline-13 comporte moins de causes d'erreur, celle-ci est considérée plus fiable.

Pour les essais suivants, le mutant D75N a été choisi pour représenter l'état inactif de hAPJ et confirmer que son système à l'équilibre n'est pas défini par une dynamique dans les mêmes régions que le système hAPJ-apeline. Le mutant N112G a été choisi pour représenter l'état actif du récepteur et confirmer que son système à l'équilibre est défini par une dynamique dans les mêmes régions que le système hAPJ-apeline. Finalement, les mutants D126V et V254L ont été choisis pour déterminer les propriétés du ou des états de hAPJ menant à une activité constitutive de $G\alpha_{i1}$.

Analyse théorique du système hAPJ-D75N-apo

Le système hAPJ-D75N-apo présente 16 états et huit composantes indépendantes. Les graphiques d'analyse par sous-matrices principales du système se trouvent à l'Annexe V.





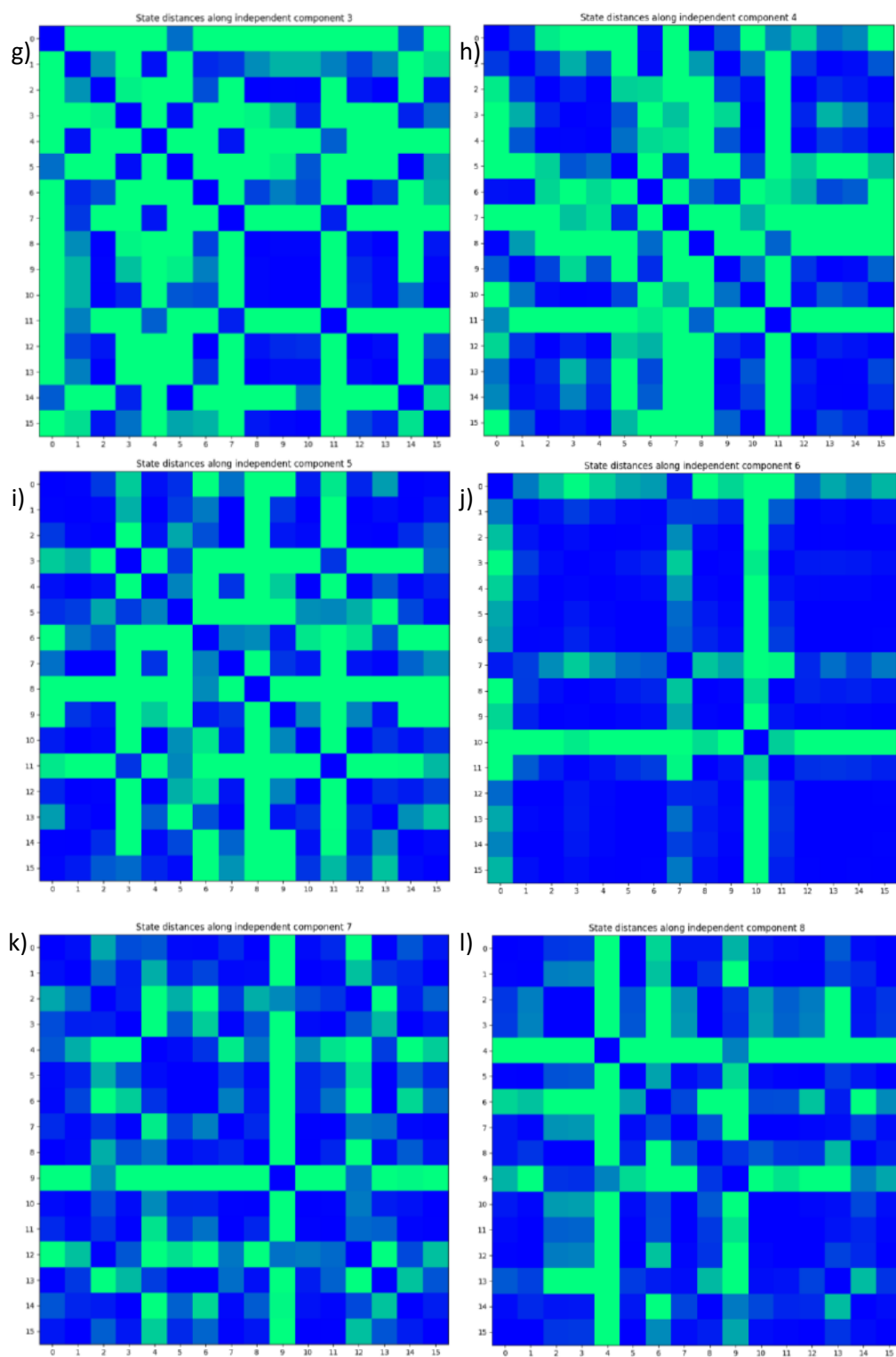
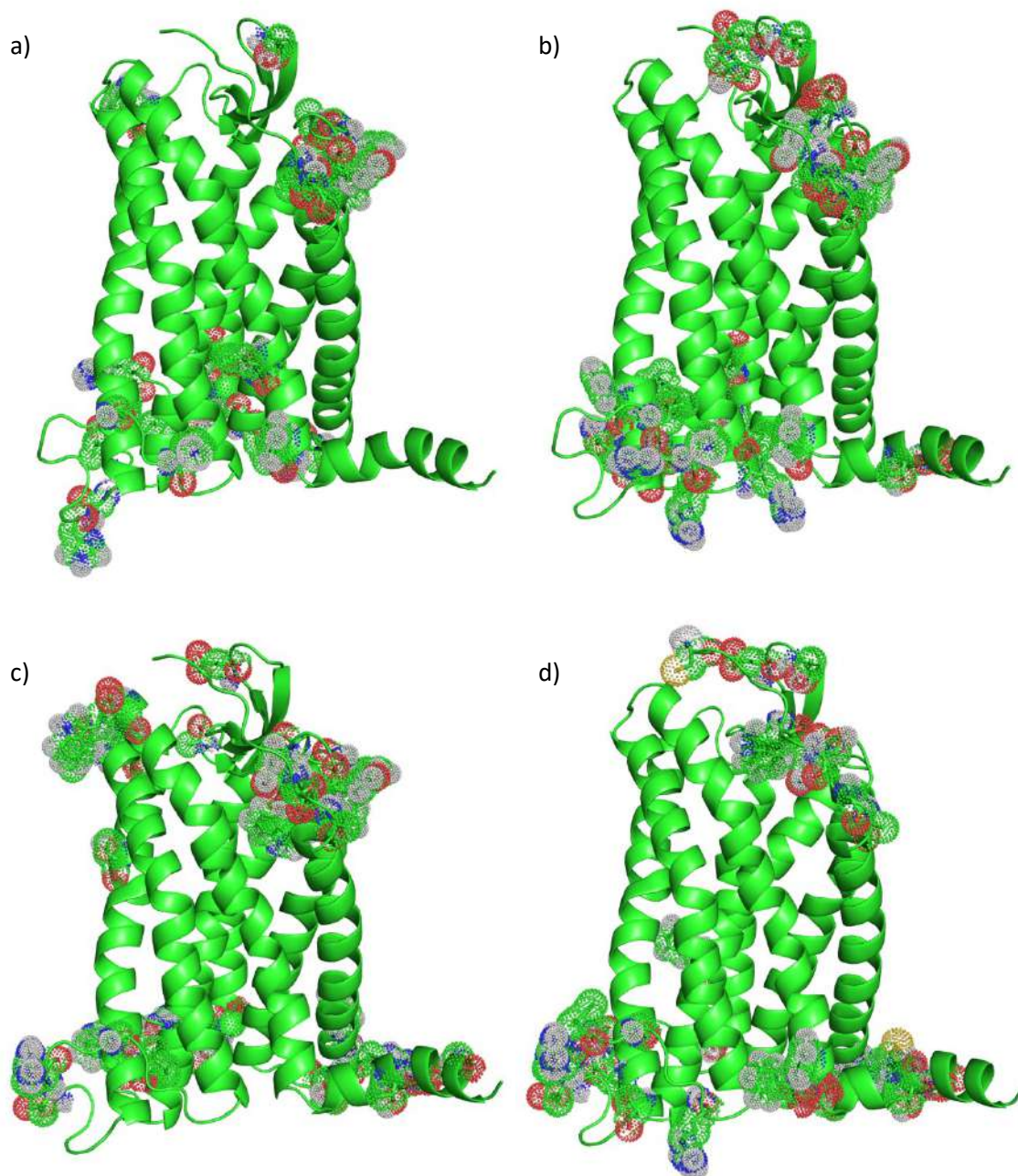


Figure 24 : Résultats du programme d'analyse de base du système hAPJ-D75N-apo
a-d) Illustration de la distribution des conformations explorées par l'ensemble des trajectoires du système hAPJ-D75N-apo. Huit composantes indépendantes et seize états (numérotés de 0 à 15) sont nécessaires pour bien décrire le système. e-l) Représentation en *heat map* de la distance de Bhattacharyya entre les différents états le long des différentes

composantes indépendantes. Les couleurs vont de bleu (0) à vert (plus de 1,982, ou erreur de classification inférieure à 2,5%). Le seul ensemble conformationnel dont la distinction du reste du système ne requiert qu'une composante est celui contenant l'état 5 (composante 2). Les autres états sont définis par un cumul des différents changements conformationnels représentés par les composantes indépendantes du système.



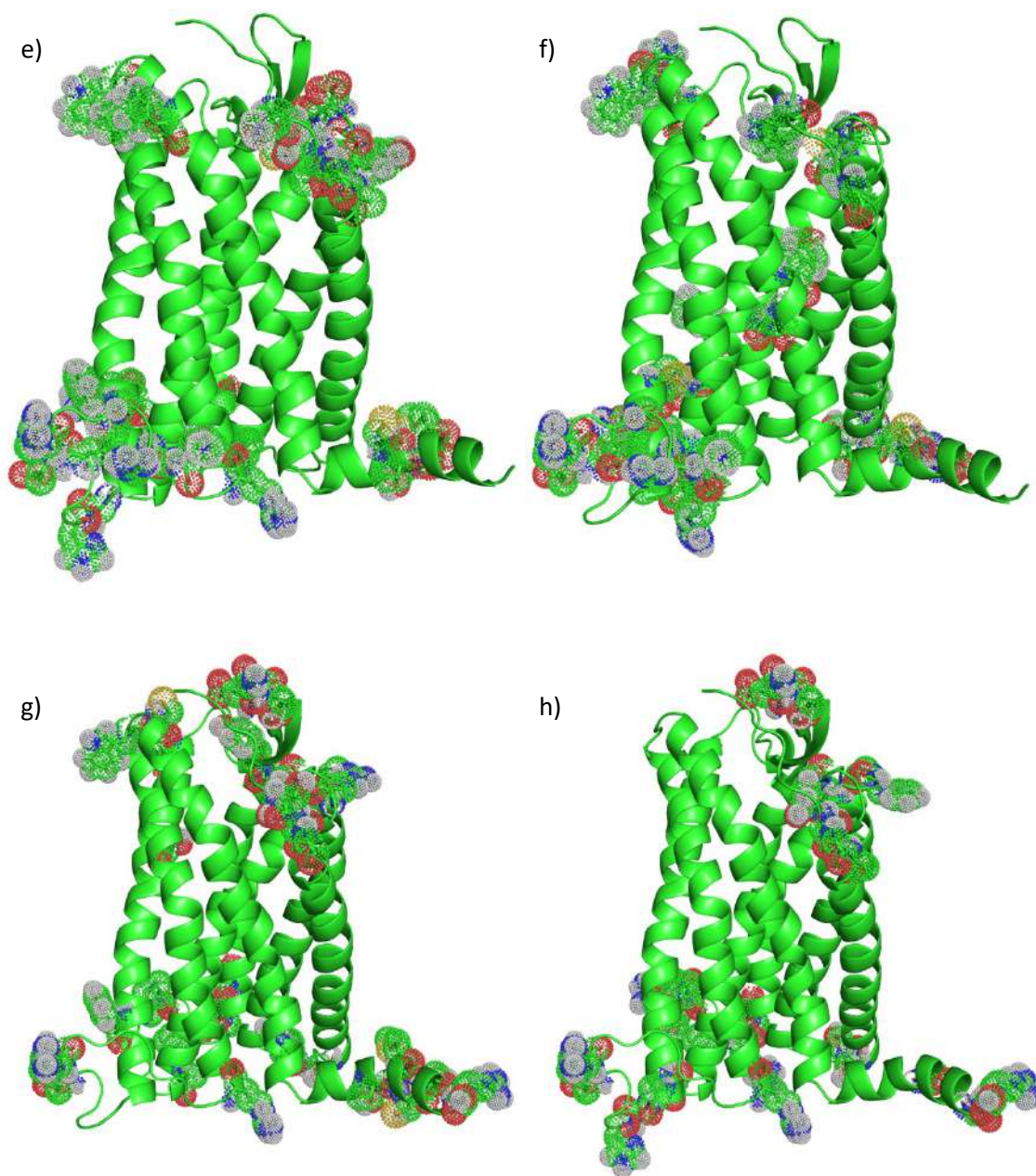
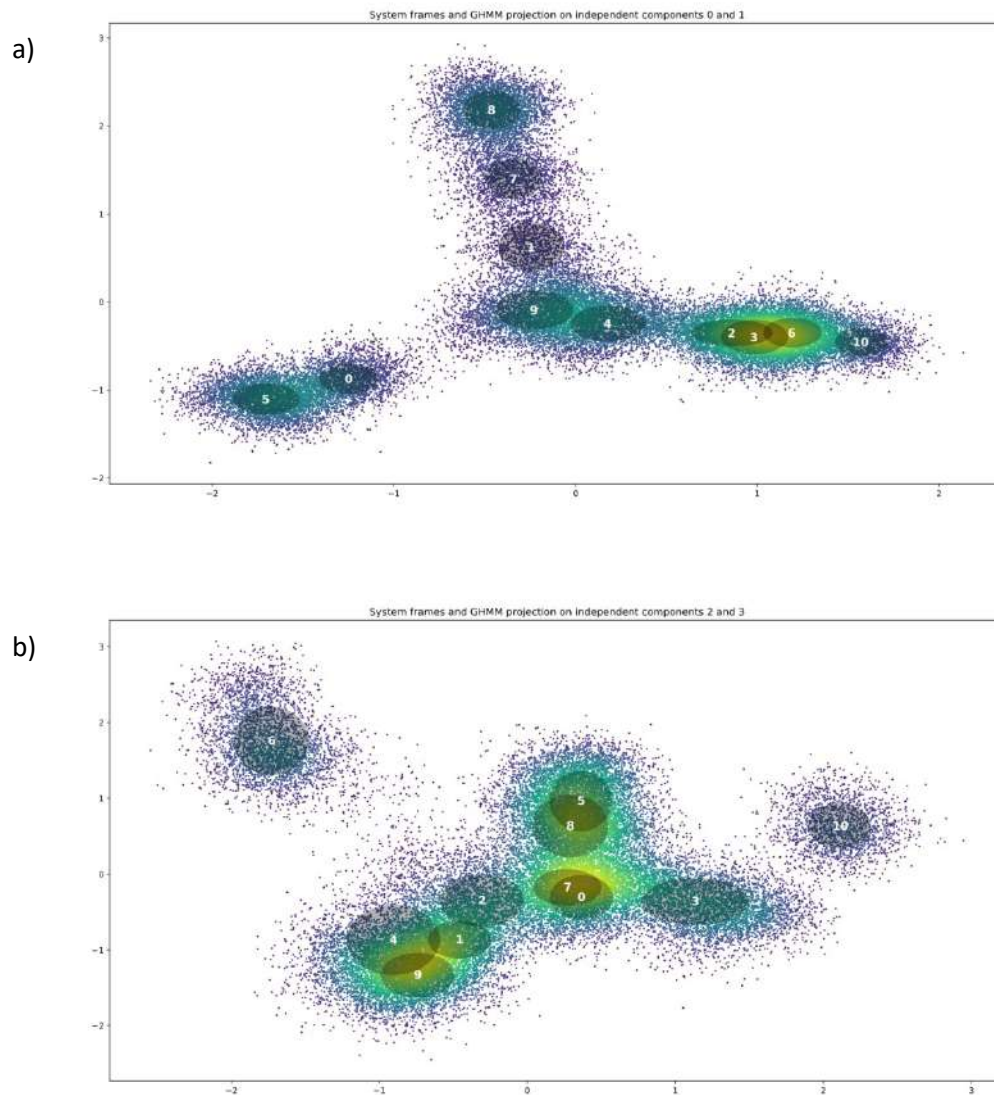


Figure 25 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-D75N-apo
a-h) Répartition sur hAPJ des résidus critiques des composantes indépendantes 1 à 8 indiqués par l'analyse par sous-matrices principales du système hAPJ-D75N-apo. Les seuls résidus impliqués hors des régions intracellulaires et intracellulaires sont des résidus de l'hélice 8 (composantes 2-8), des résidus autour du site de la mutation (composantes 1 et 6), le résidu F118^{3,41} (composantes 4 et 6) et les résidus T206^{5,43} et S205^{5,42} (composantes 3 et 7, respectivement). Les différents états du système témoignent d'une distorsion formant un « coude » dans le TM3 au niveau de F118, d'une distorsion du TM5 au niveau de S205 et T206 et de réarrangements substantiels au niveau de l'hélice 8.

Analyse théorique du système hAPJ-N112G-apo

Le système hAPJ-N112G-apo présente quatre composantes indépendantes et onze états. Les graphiques d'analyse par sous-matrices principales du système se trouvent à l'Annexe VI.



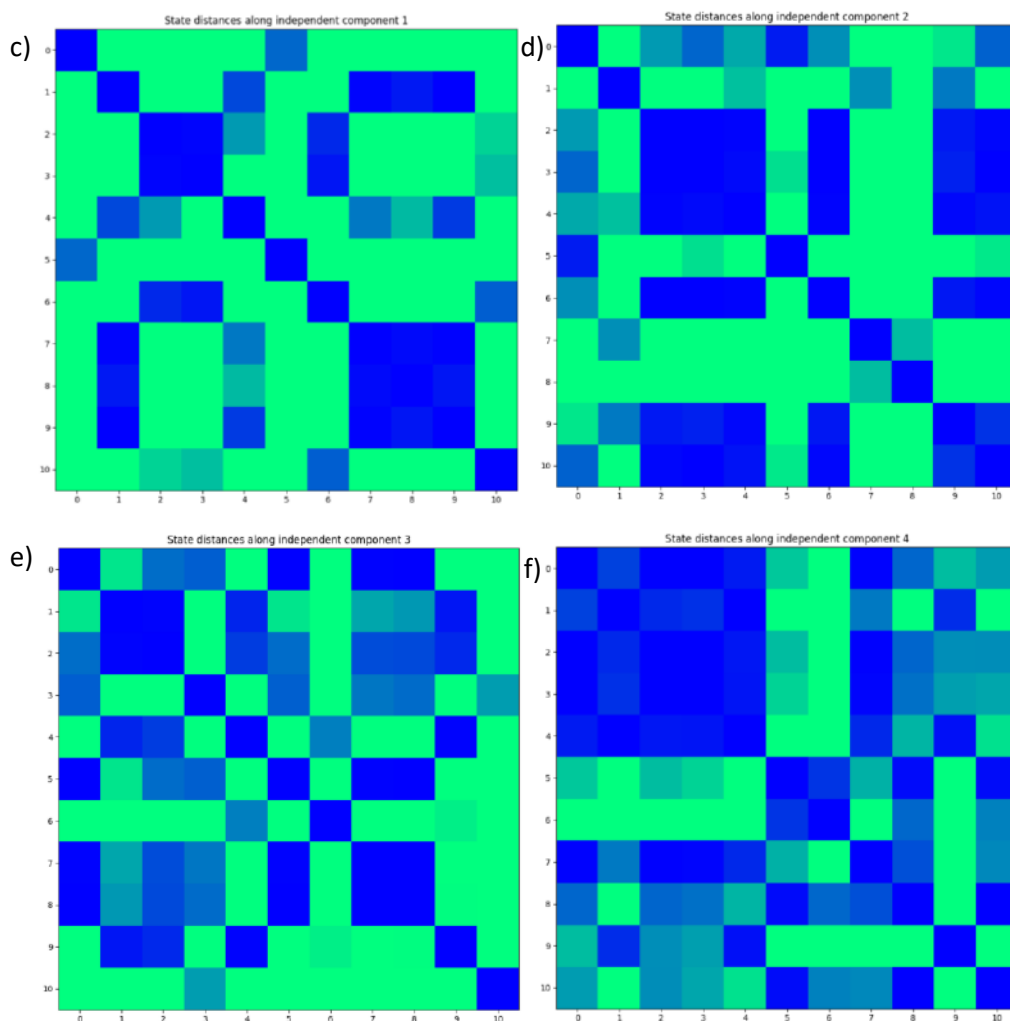


Figure 26 : Résultats du programme d'analyse de base du système hAPJ-N112G-apo

a, b) Illustration de la distribution des conformations explorées par l'ensemble des trajectoires du système hAPJ-N112G-apo. Quatre composantes indépendantes et onze états (numérotés de 0 à 10) sont nécessaires pour bien décrire le système. c-f) Représentation en *heat map* de la distance de Bhattacharyya entre les différents états le long des différentes composantes indépendantes. Les couleurs vont de bleu (0) à vert (plus de 1,982, ou erreur de classification inférieure à 2,5%). Le seul ensemble conformationnel dont la distinction du reste du système ne requiert qu'une composante est celui contenant les états 0 et 5 (composante 1). Les autres états sont définis par un cumul des différents changements conformationnels représentés par les composantes indépendantes du système.

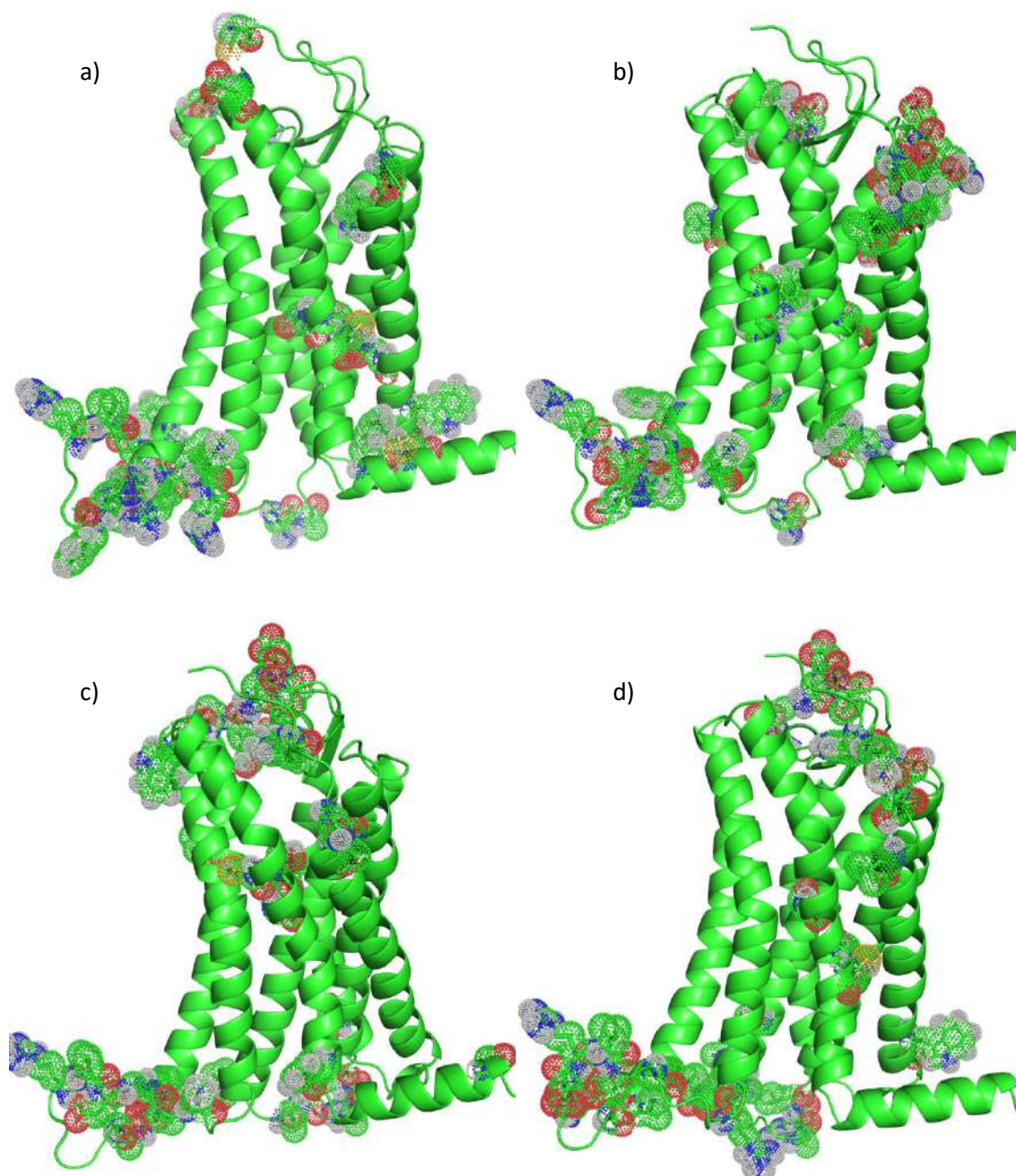
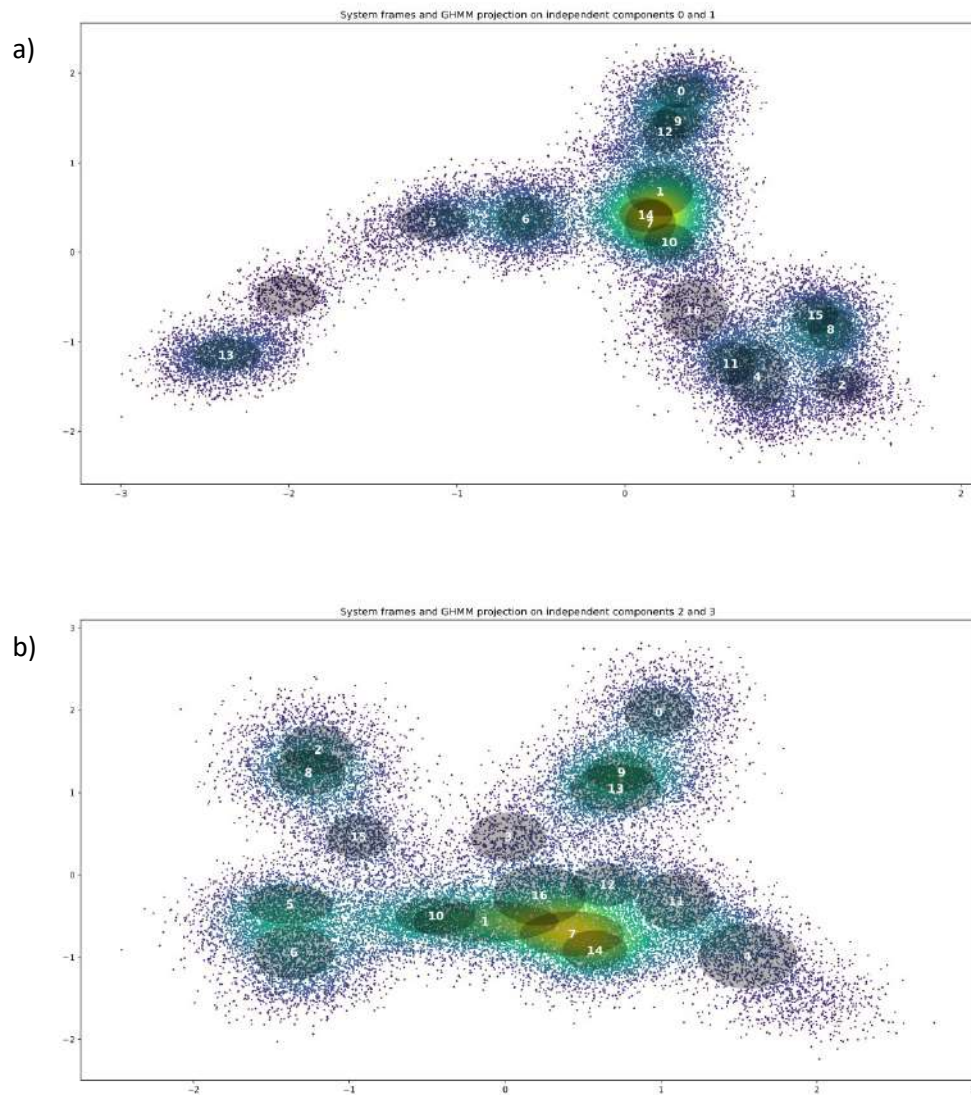
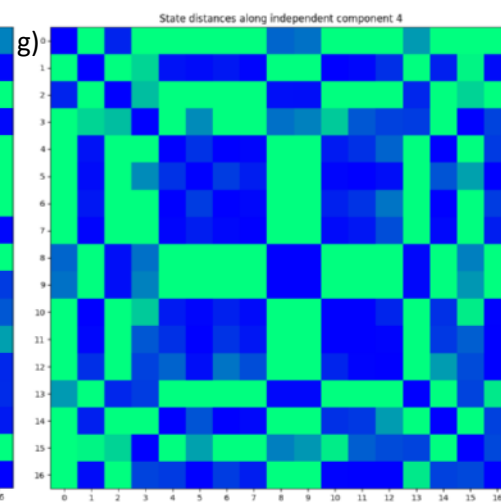
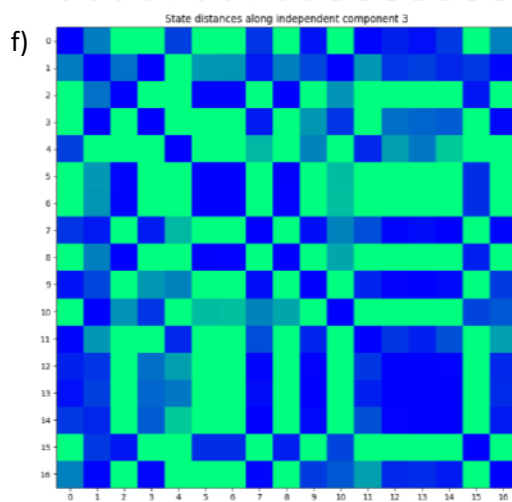
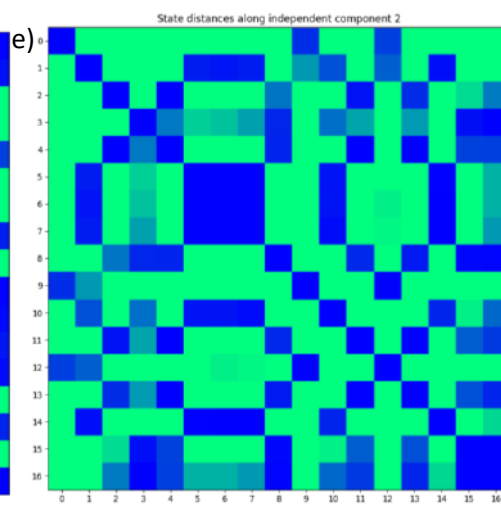
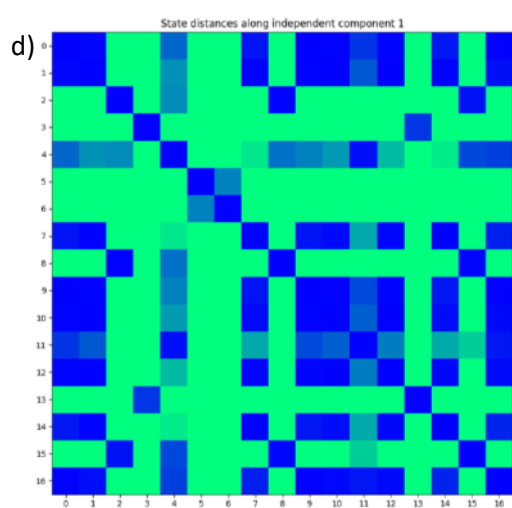
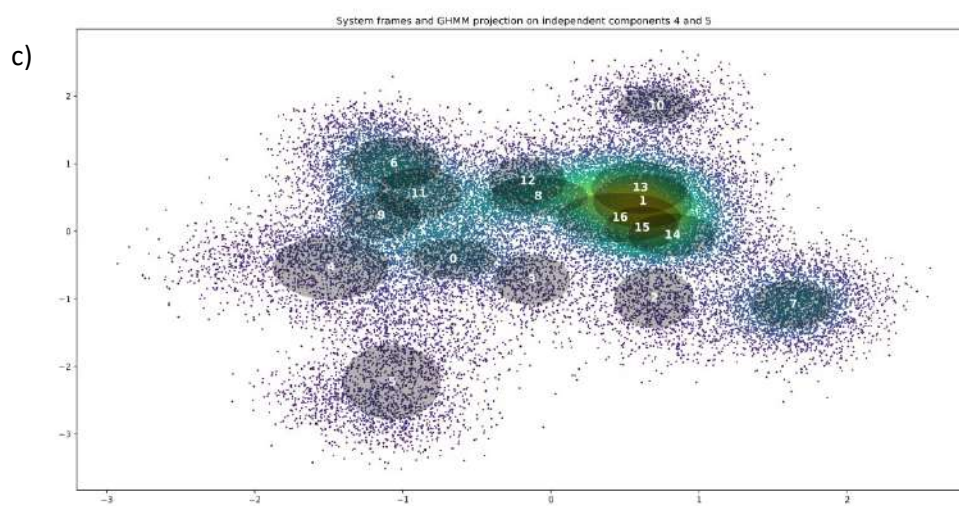


Figure 27 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-N112G-apo
a-d) Répartition sur hAPJ des résidus critiques des composantes indépendantes 1 à 4 indiqués par l'analyse par sous-matrices principales du système hAPJ-N112G-apo. À l'exception des boucles intracellulaires et extracellulaires, les régions impliquées dans les changements conformationnels représenté par ces composantes se concentrent au niveau du HHM et du MHBN, avec des signatures similaires, mais moins étendues que celles du système hAPJ-apeline.

Analyse théorique du système hAPJ-D126V-apo

Le système hAPJ-D126V-apo présente six composantes indépendantes et 17 états, numérotés de 0 à 16. Les graphiques d'analyse par sous-matrices principales du système se trouvent à l'Annexe VII.





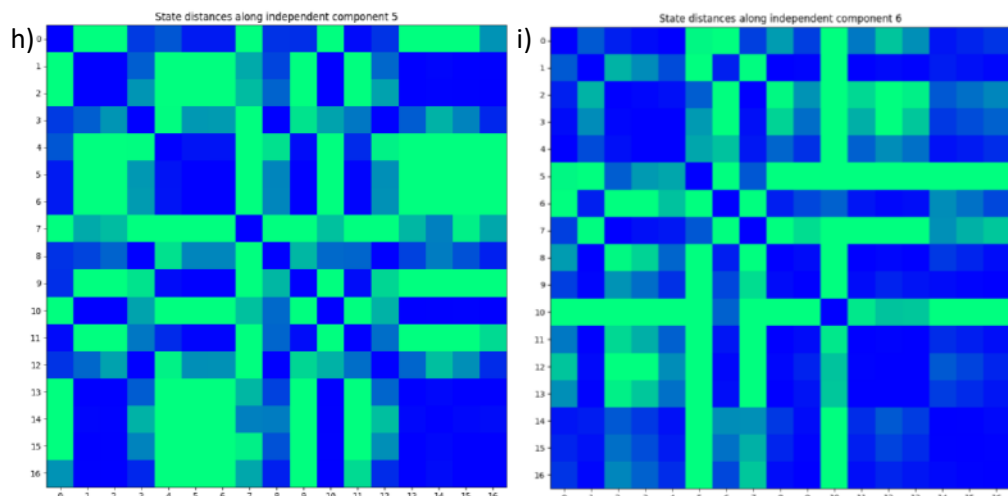
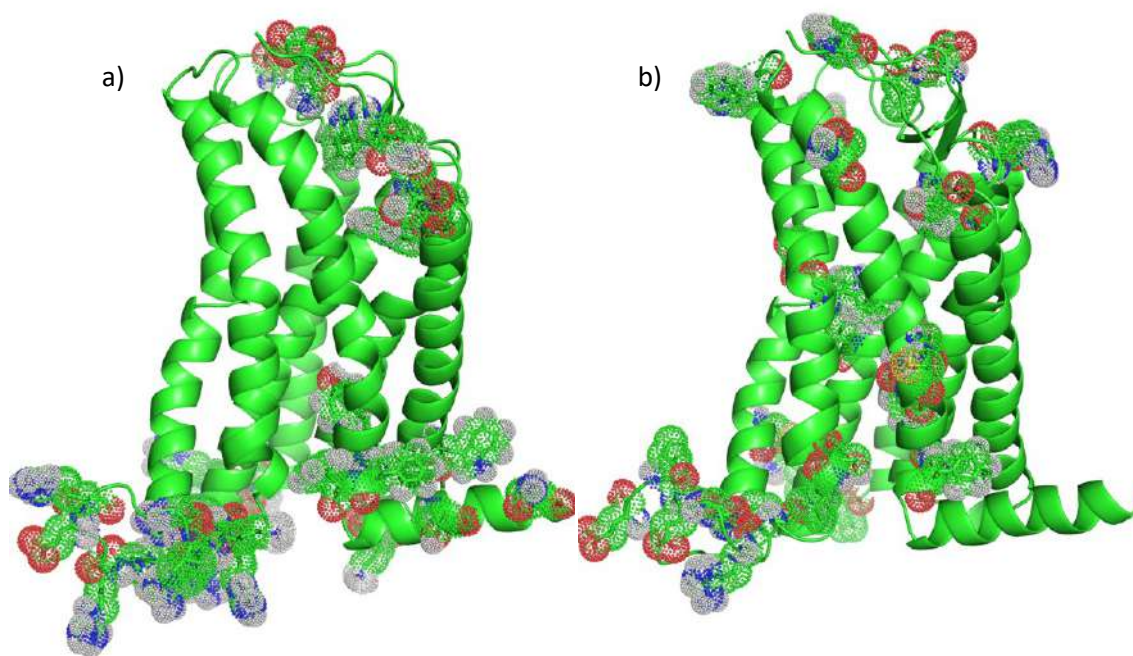


Figure 28 : Résultats du programme d'analyse de base du système hAPJ-D126V-apo

a-c) Illustration de la distribution des conformations explorées par l'ensemble des trajectoires du système hAPJ-D126V-apo. Six composantes indépendantes et dix-sept états (numérotés de 0 à 16) sont nécessaires pour bien décrire le système. d-i) Représentation en *heat map* de la distance de Bhattacharyya entre les différents états le long des différentes composantes indépendantes. Les couleurs vont de bleu (0) à vert (plus de 1,982, ou erreur de classification inférieure à 2,5%). Les ensembles conformationnels dont la distinction du reste du système ne requiert qu'une composante sont ceux contenant les états 3 et 13, ainsi que 5 et 6 (composante 1). Les autres états sont définis par un cumul des différents changements conformationnels représentés par les composantes indépendantes du système.



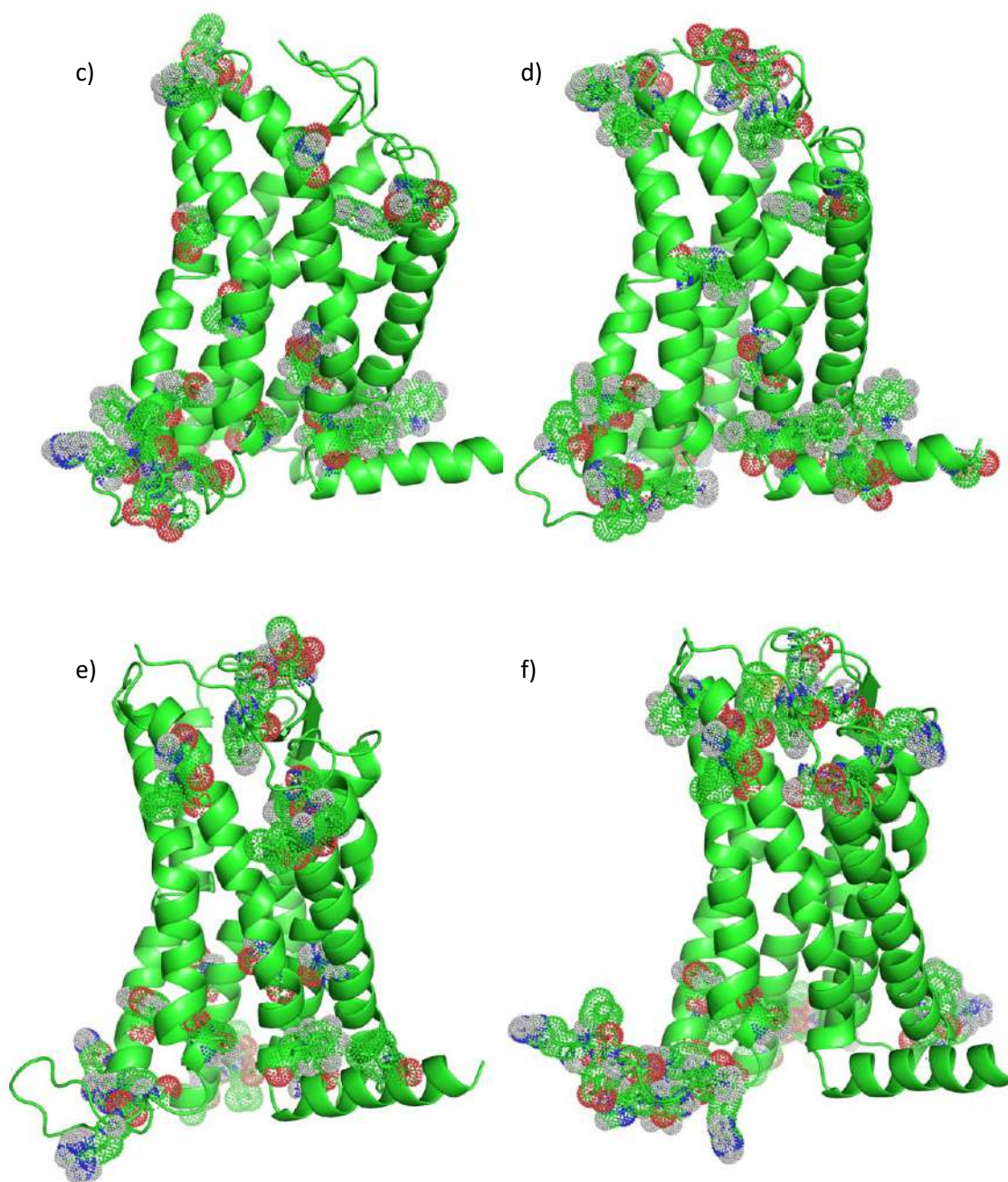
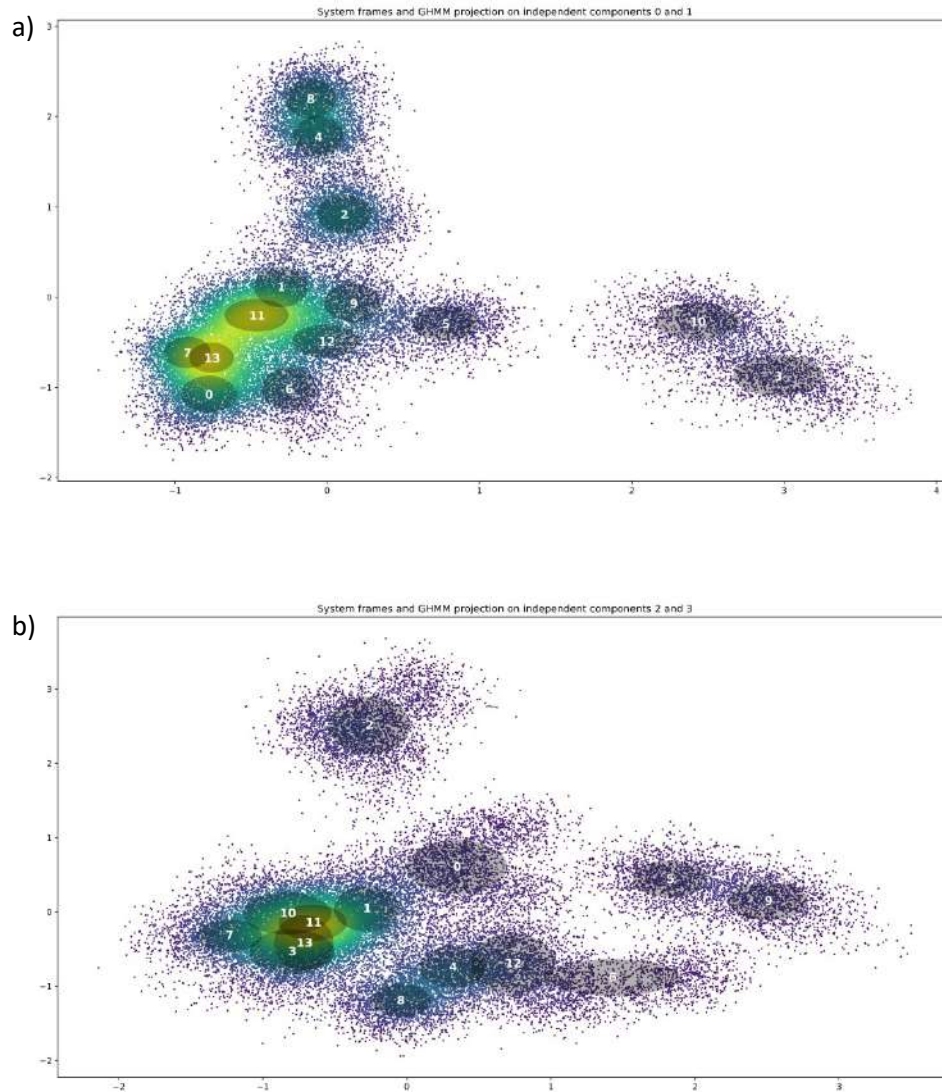


Figure 29 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-D126V-apo

a-f) Répartition sur hAPJ des résidus critiques des composantes indépendantes 1 à 6 indiqués par l'analyse par sous-matrices principales du système hAPJ-D126V-apo. La dynamique du système implique un réarrangement de l'interface entre les TM3, 5 et 6, ainsi que du TM7 au niveau intracellulaire. En effet, les changements conformationnels représentés par les six composantes impliquent les résidus V117^{3,40} (composante 2), T206^{5,43} (composantes 2 et 3), C220^{5,57} (composante 2), Y221^{5,58} (composantes 3 et 4), F222^{5,60} (composante 4), F223^{5,61} (composante 3), A268^{6,45} (composante 3), W261^{6,48} (composantes 2 et 4) et les résidus du motif NPxxY, plus ou moins un tour d'hélice (composantes 1, 2, 3 et 4).

Analyse théorique du système hAPJ-V254L-apo

Le système hAPJ-V254L-apo présente quatre composantes indépendantes et quatorze états. Les graphiques d'analyse par sous-matrices principales du système se trouvent à l'Annexe VIII.



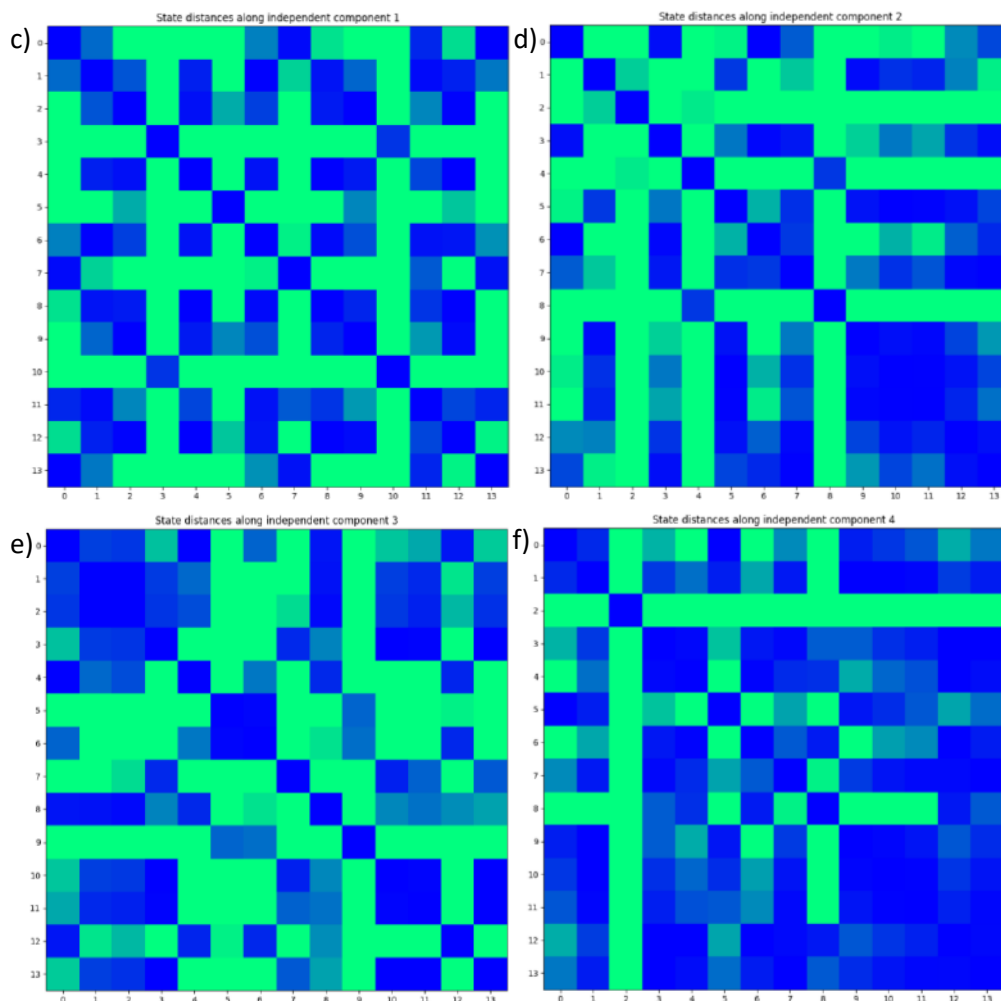


Figure 30 : Résultats du programme d'analyse de base du système hAPJ-V254L-apo

a, b) Illustration de la distribution des conformations explorées par l'ensemble des trajectoires du système hAPJ-V254L-apo. Quatre composantes indépendantes et quatorze états (numérotés de 0 à 13) sont nécessaires pour bien décrire le système. c-f) Représentation en *heat map* de la distance de Bhattacharyya entre les différents états le long des différentes composantes indépendantes. Les couleurs vont de bleu (0) à vert (plus de 1,982, ou erreur de classification inférieure à 2,5%). Les ensembles conformationnels dont la distinction du reste du système ne requiert qu'une composante sont ceux contenant les états 3 et 10 (composante 1) et l'état 2 (composante 4). Les autres états sont définis par un cumul des différents changements conformationnels représentés par les composantes indépendantes du système.

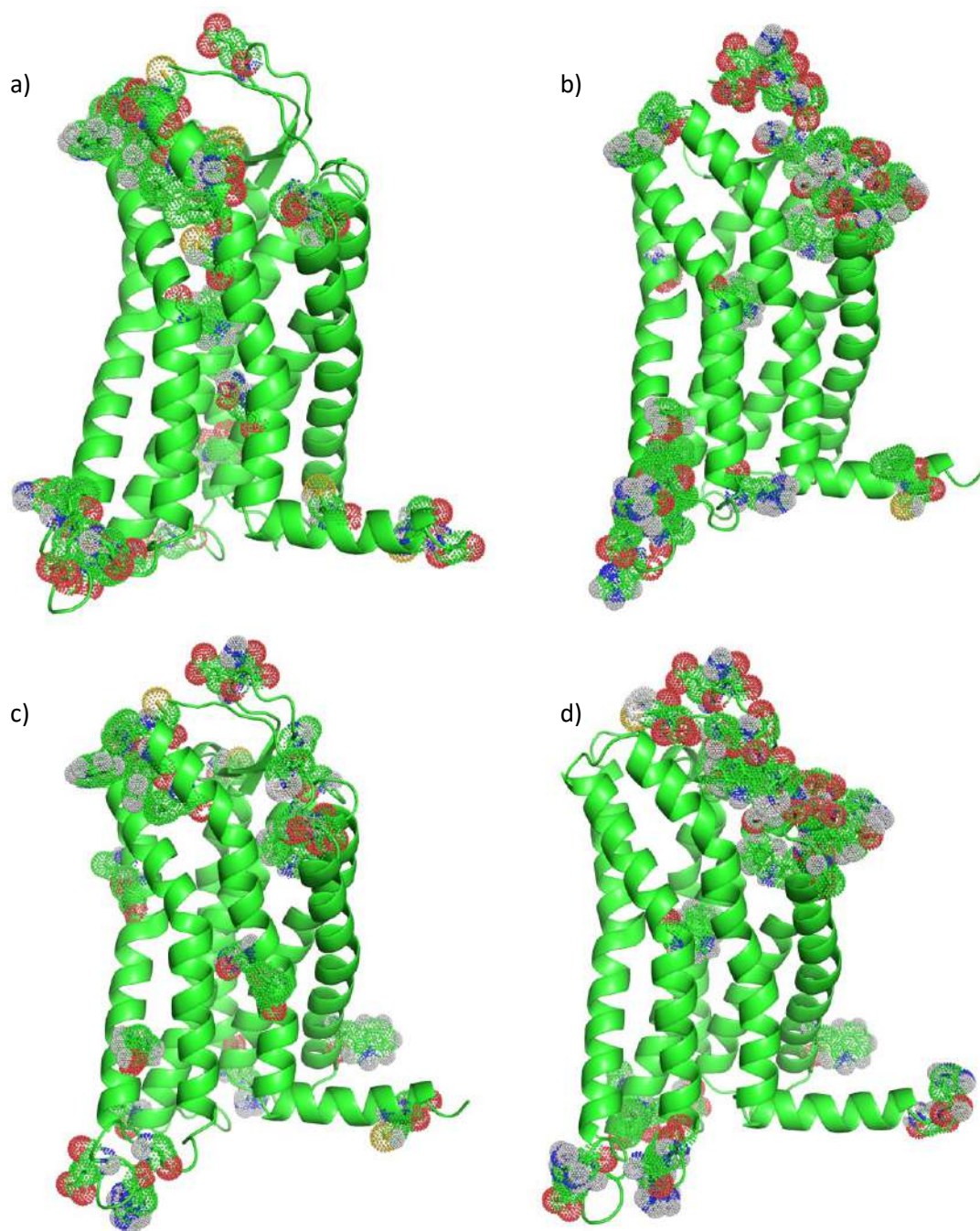
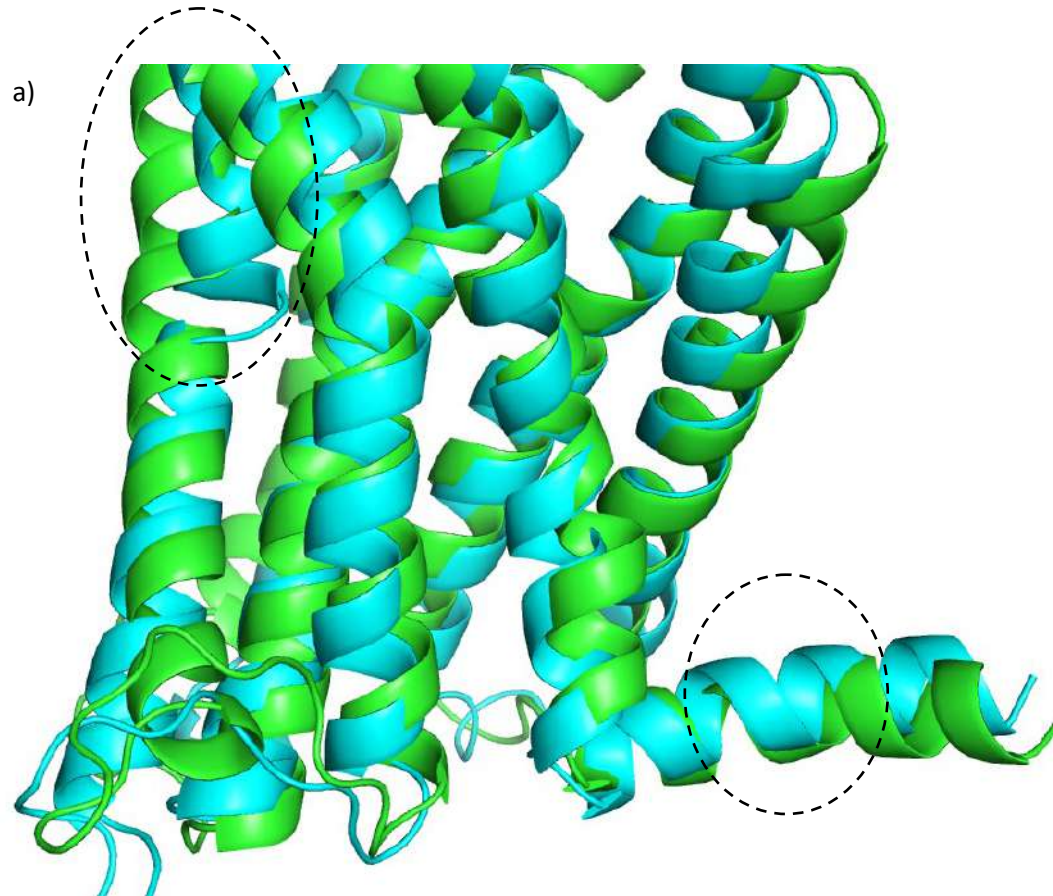


Figure 31 : Analyse conformationnelle des résultats de l'analyse du système hAPJ-V254L-apo
a-d) Répartition sur hAPJ des résidus critiques des composantes indépendantes 1 à 4 indiqués par l'analyse par sous-matrices principales du système hAPJ-V254L-apo. La dynamique du système implique un réarrangement de l'interface entre les TM5 et 6, ainsi que du TM7 au niveau intracellulaire. En effet, les changements conformationnels représentés par les quatre composantes impliquent les résidus V203^{5,40}, S204^{5,41}, T206^{5,43}, Y221^{5,58} (Composantes 2 et 3), W261^{6,48} (Composantes 1, 2 et 4), les quatre premiers tours d'hélice du TM7 (Composantes 1 et 3), le tour d'hélice au-dessus du motif NPxxY (Composante 3).

Analyse comparative théorique des systèmes mutants

Comparaison des systèmes hAPJ-WT-apo et hAPJ-D75N-apo

Les deux systèmes diffèrent principalement à trois endroits : la portion extracellulaire du TM5, l'interface entre les TM2, 3 et 4, ainsi que l'hélice 8.



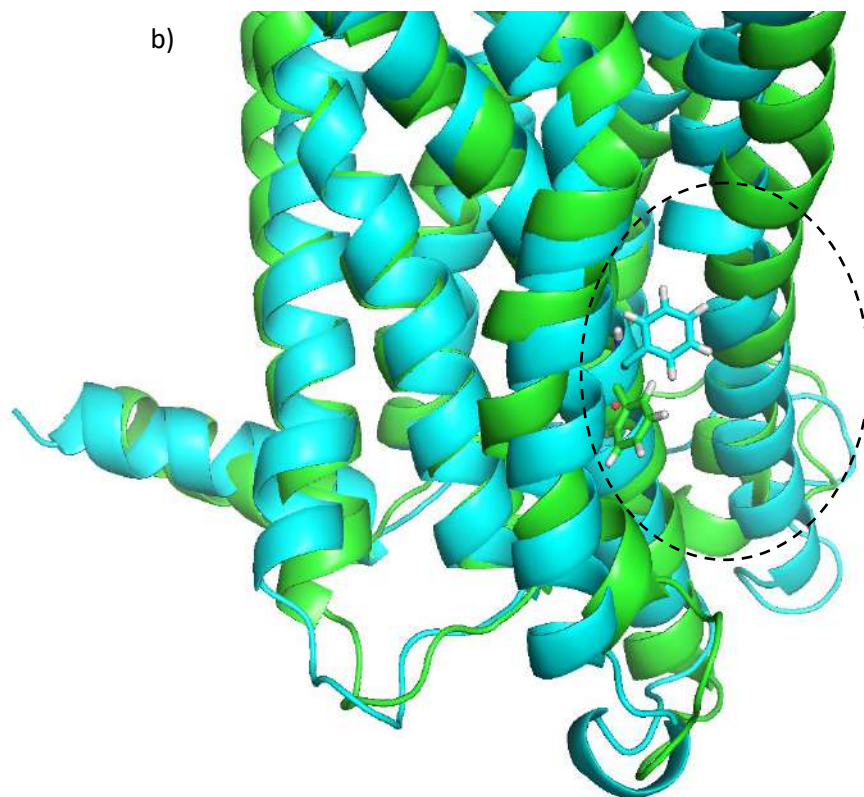
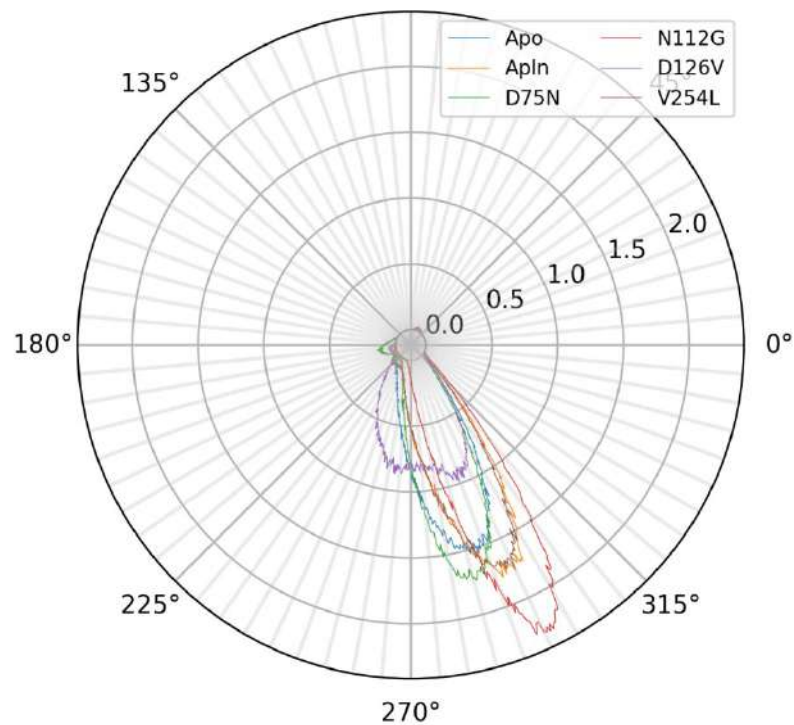


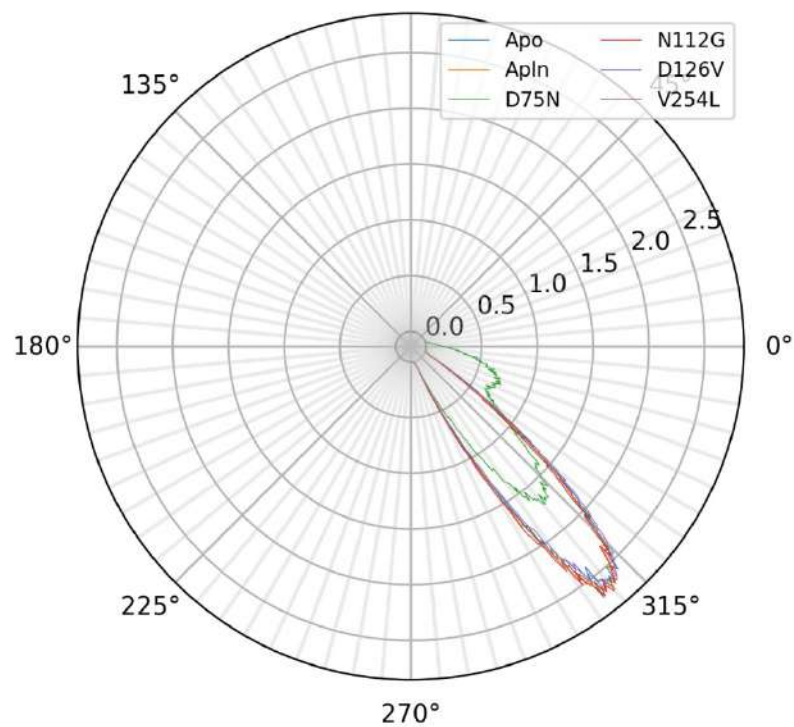
Figure 32 : Différences entre deux conformations choisies des systèmes hAPJ-WT-apo et hAPJ-D75N-apo

a) Identification des régions du TM5 et de l'hélice 8 les plus différentes entre le système hAPJ-WT-apo (bleu, état 11) et le système hAPJ-D75N-apo (vert, état 12). b) Illustration du changement conformationnel affectant l'interface des TM2, 3 et 4. F118 change de conformation (χ_1) et provoque une distorsion du TM4 au niveau de A149 (ψ), ce qui le rapproche du TM2.

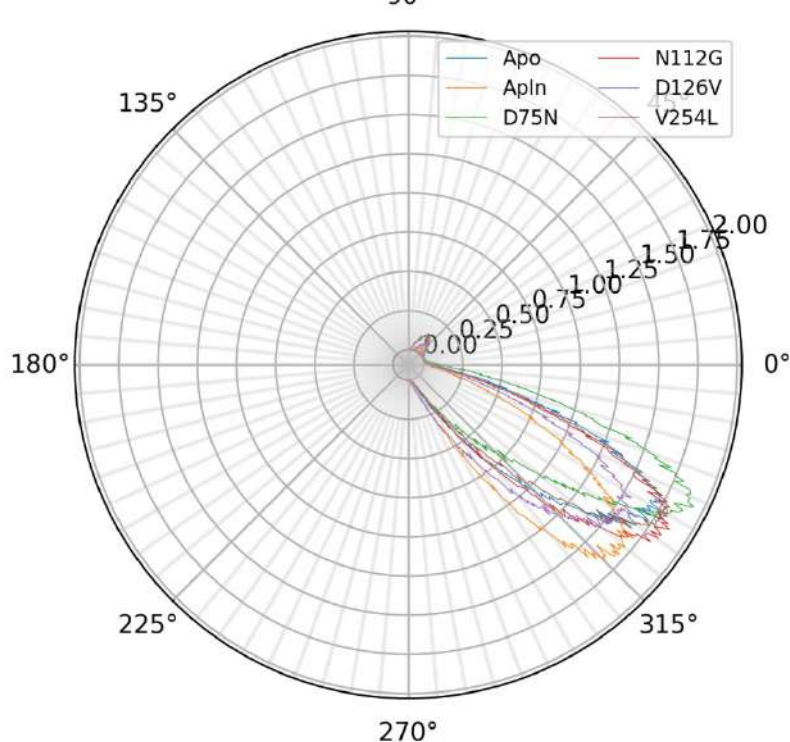
a) Distribution of PHE118- χ_1 angle for all hAPJ systems



b) Distribution of ALA149- ψ angle for all hAPJ systems



c) Distribution of VAL208- ψ angle for all hAPJ systems



d) Distribution of CYS320- ψ angle for all hAPJ systems

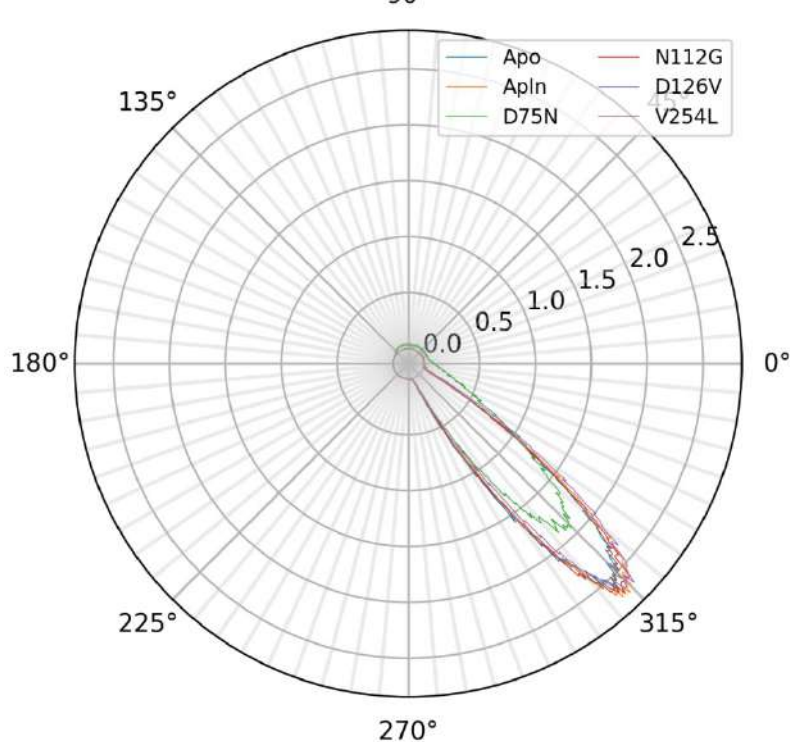
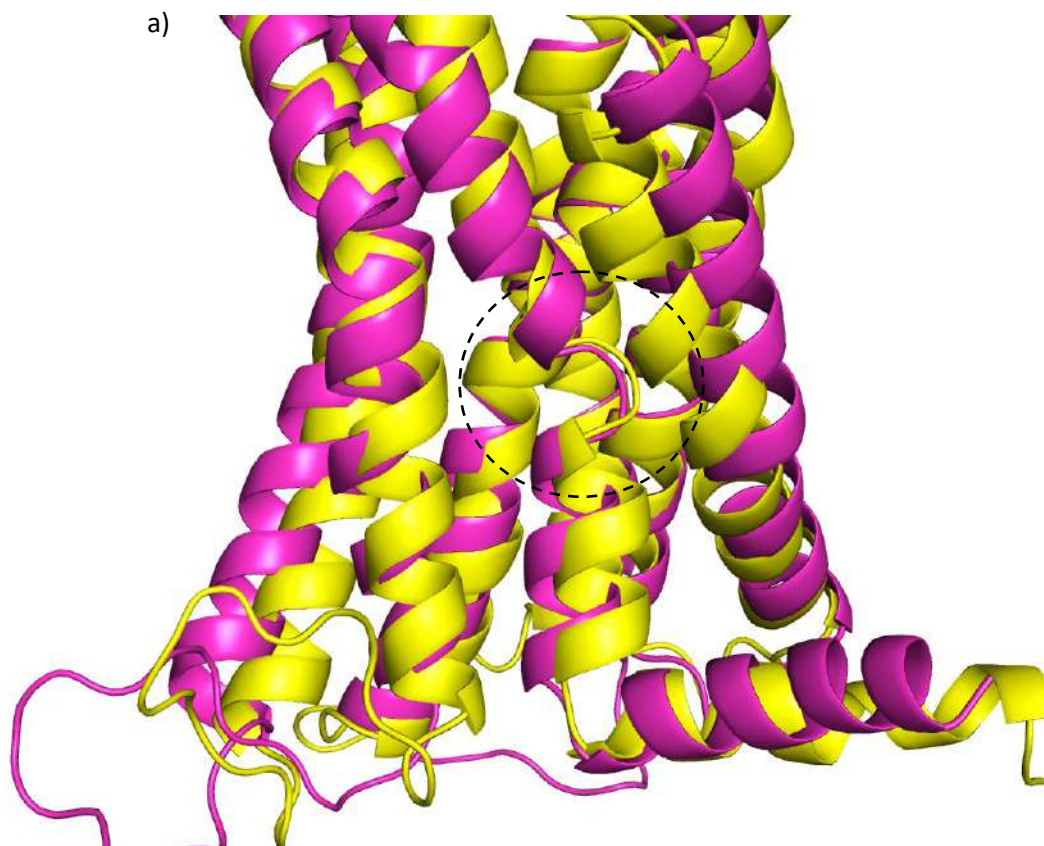


Figure 33 : Distributions d'angles dièdres d'intérêt pour la comparaison des systèmes hAPJ-WT-apo et hAPJ-D75N-apo

Le changement conformationnel dans la région de l'interface des TM2, 3 et 4 s'explique principalement par la variation des angles dièdres a) F118- χ_1 et b) A149- ψ . En a), il est possible de voir que l'angle F118- χ_1 varie de 100 degrés dans le système hAPJ-D75N-apo, passant de 285 à 185°. En b), il est possible de voir que l'angle A149- ψ est sujet à des variations importantes dans le système hAPJ-D75N-apo, contrairement aux autres systèmes. La différence de conformation de la partie extracellulaire du TM5 est explicable par la différence des angles du squelette de l'hélice α autour de V208. Certaines de ces variations, comme celle de c) V208- ψ , sont uniquement des variations de la moyenne des angles, pas des changements conformationnels, donc invisibles du point de vue de la tICA. Il est à noter que des changements mineurs dans la valeur d'angles dièdres du squelette de la protéine, en particulier dans le cas de longues hélices α , a un impact significatif sur la conformation de cette hélice par rapport au reste de la protéine. Les variations de la conformation de l'hélice 8 s'expliquent bien par les variations des angles dièdres du squelette de la protéine autour de C208. En d), on peut observer que l'angle C208- ψ , dans l'hélice 8, possède une distribution bien plus étendue pour le système hAPJ-D75N-apo que pour les autres systèmes.

Comparaison des systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ-N112G-apo

Les changements conformationnels observés dans les systèmes hAPJ-WT-apeline et hAPJ-N112G-apo sont similaires. Tous deux se produisent au niveau de F257 et du tour d'hélice au-dessus du motif NPxxY.



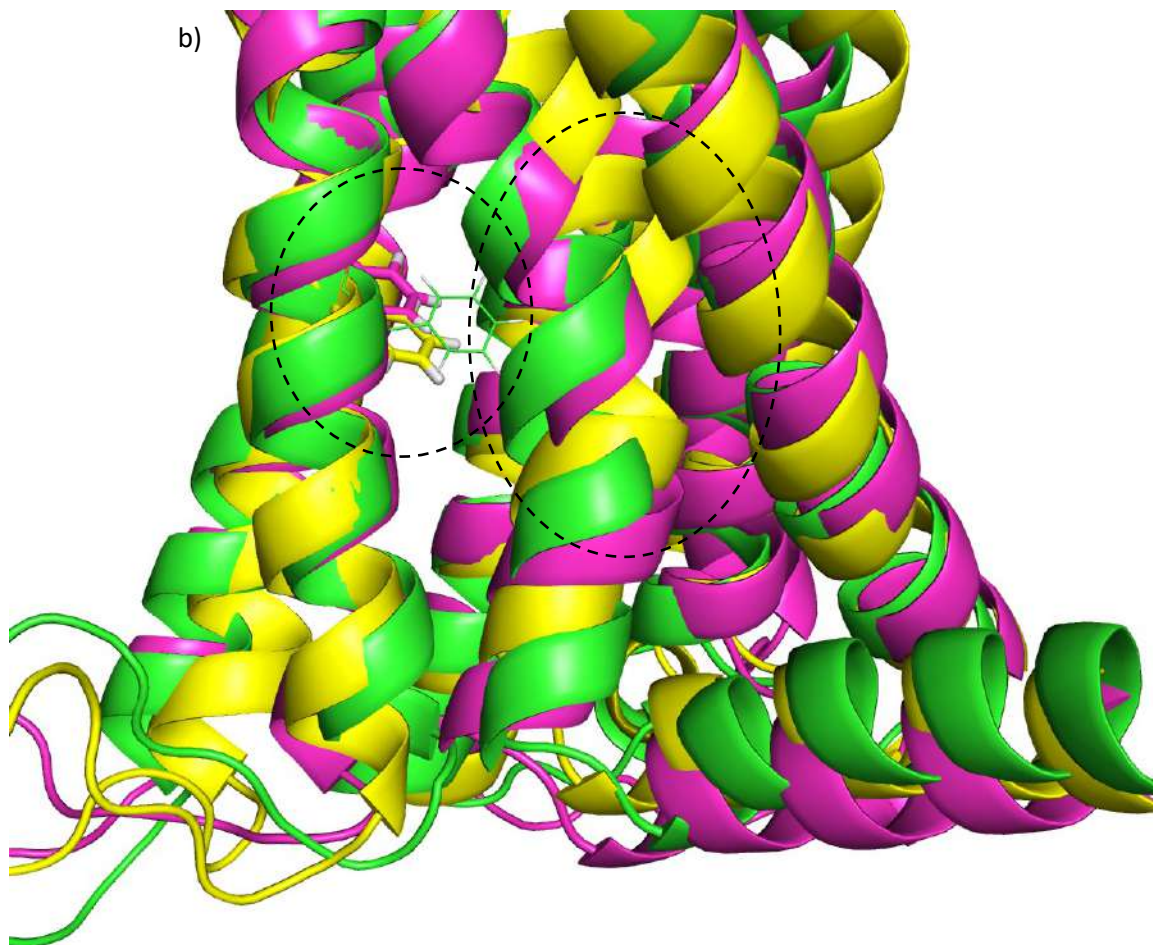
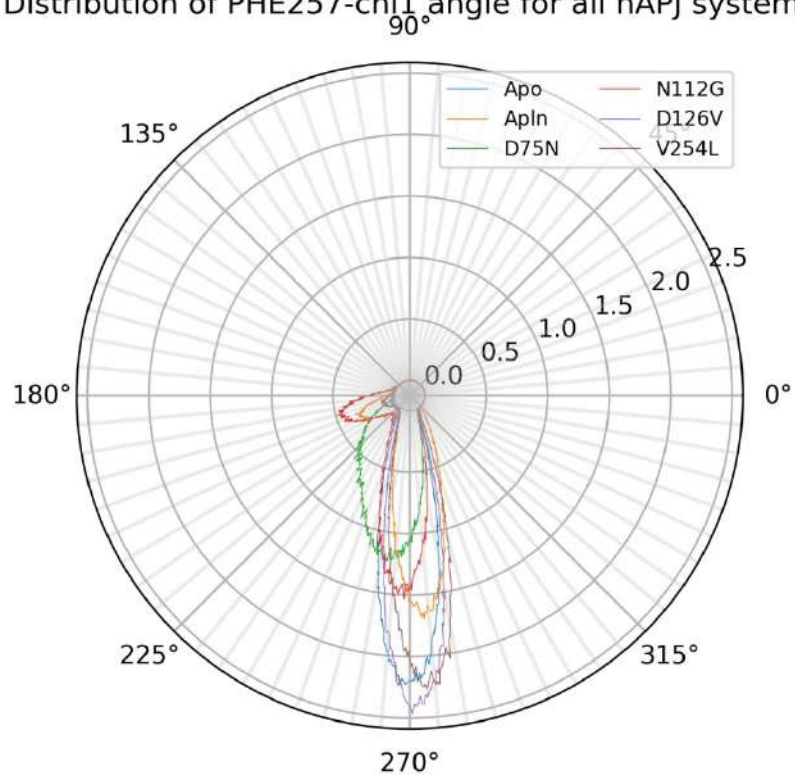


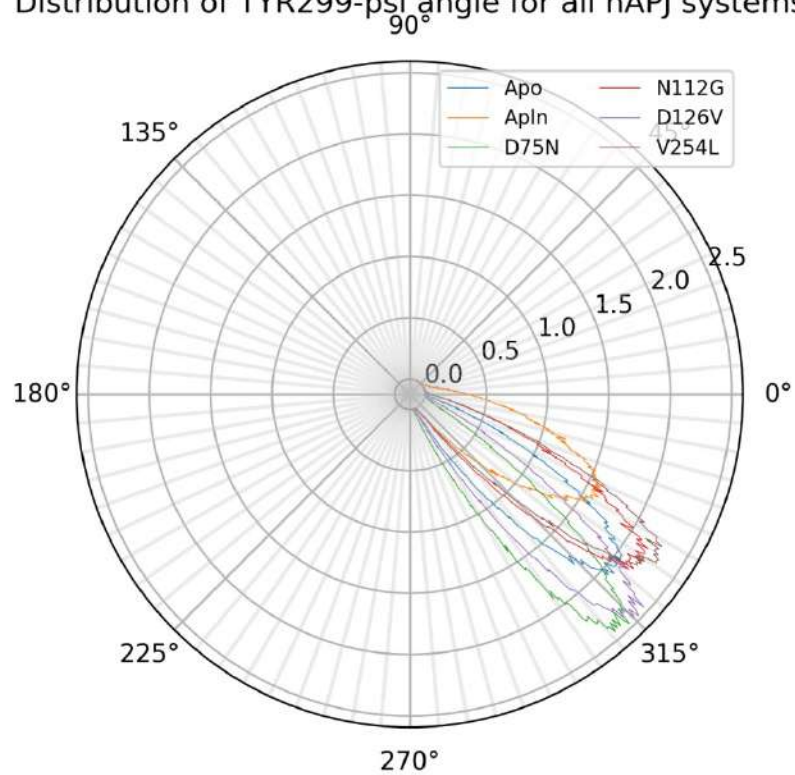
Figure 34 : Illustration des deux changements conformationnels majeurs observés seulement dans les systèmes hAPJ-WT-apeline et hAPJ-N112G-apo

a) Illustration de la dénaturation locale de l'hélice α dans le tour d'hélice situé au-dessus du motif NPxxY. Autant le système hAPJ-WT-apeline (jaune, état 7) que le système hAPJ-N112G-apo (magenta, état 5) ont des états dont les images présentent ce changement conformationnel. b) Illustration du changement conformationnel de F257 et la distorsion correspondante du TM7 au-dessus du motif NPxxY des systèmes hAPJ-WT-apeline (jaune, état 0) et hAPJ-N112G-apo (magenta, état 7). En guise de comparaison, le système hAPJ-WT-apo (vert, état 11) ne présente pas le changement conformationnel observé sur F257. De plus, le TM7 de ce système est relativement droit par rapport aux deux autres.

a) Distribution of PHE257- χ_1 angle for all hAPJ systems



b) Distribution of TYR299- ψ_1 angle for all hAPJ systems



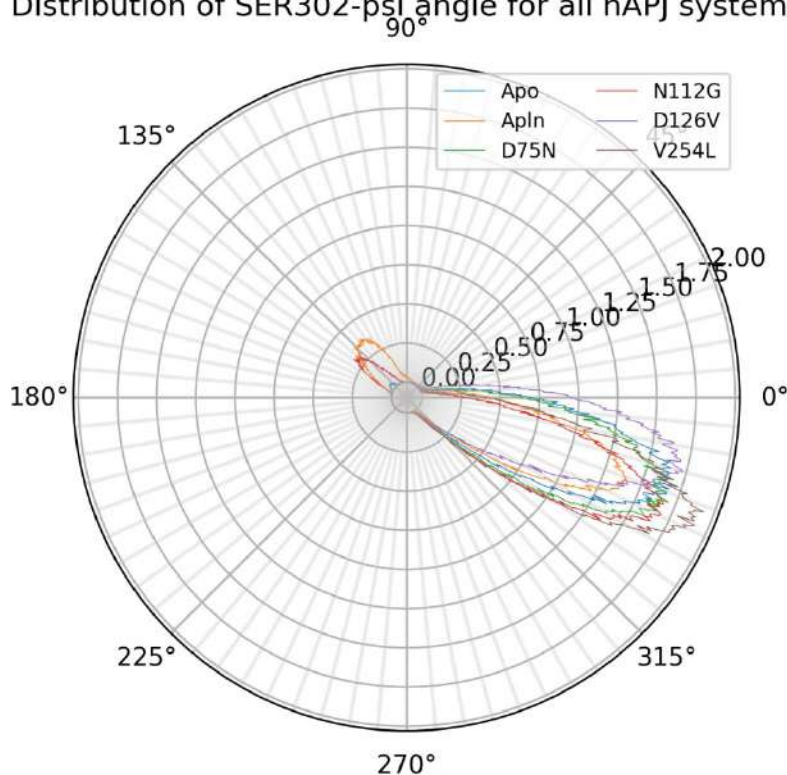
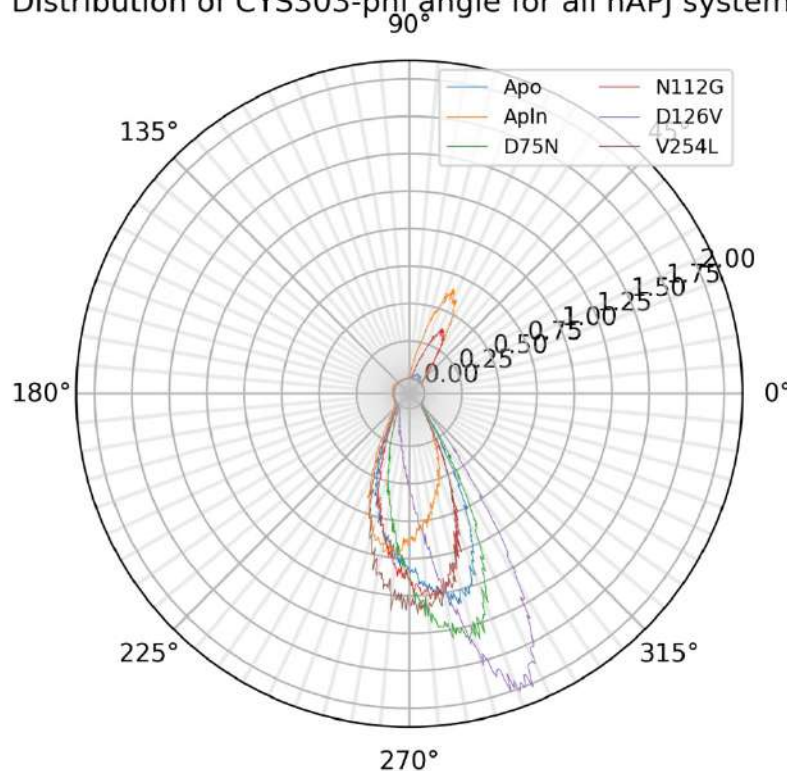
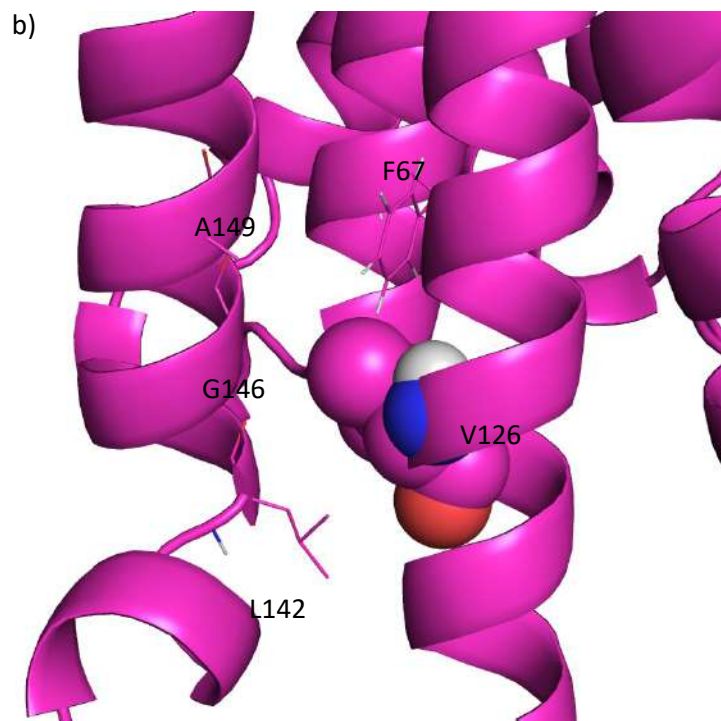
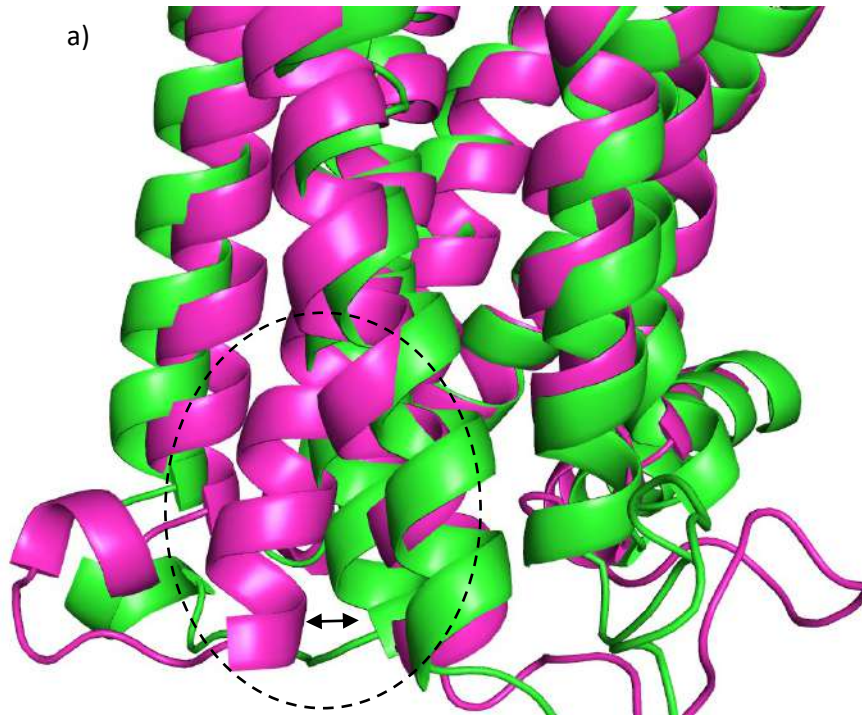
c) Distribution of SER302- ψ angle for all hAPJ systemsd) Distribution of CYS303- ϕ angle for all hAPJ systems

Figure 35 : Distributions d'angles dièdres d'intérêt pour la comparaison des systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ-N112G-apo

Le changement conformationnel du résidu F257 s'explique principalement par la variation de son angle χ_1 . En a), il est possible de voir que l'angle F257- χ_1 varie de 75 degrés dans les systèmes hAPJ-WT-apeline et hAPJ-N112G-apo, passant de 270 à 195°. La variation substantielle de cet angle dans le système hAPJ-D75N-apo est moins contrainte, mais plus restreinte en amplitude, ce qui indique que l'environnement de ce résidu a changé, probablement à cause du réarrangement du TM5 dans ce système. En b), il est possible de voir que l'angle Y299- ψ varie légèrement entre les systèmes hAPJ-WT-apo et les deux autres systèmes. Cette variation est moins prononcée dans le système hAPJ-N112G-apo (5°) que le système hAPJ-WT-apeline (10°). En c), il est possible de constater que l'angle S202- ψ , varie en moyenne pour le système hAPJ-N112G-apo (5°) et que les systèmes hAPJ-N112G-apo et hAPJ-WT-apeline ont tous deux un ou plusieurs états présentant une variation importante de cet angle, ce qui explique qu'il existe des états pour ce système où la structure hélicale est localement dénaturée. Hormis la variation de la moyenne dans le système hAPJ-N112G-apo, les mêmes conclusions peuvent être tirées pour l'angle d) C303- ϕ . De plus, en c) et en d), on peut observer que la variation de la moyenne du système hAPJ-D126V-apo est opposée à celle des systèmes hAPJ-WT-apeline et hAPJ-N112G-apo.

Comparaison des systèmes hAPJ-WT-apo et hAPJ-D126V-apo

Les systèmes hAPJ-WT-apo et hAPJ-D126V-apo diffèrent en deux endroits en particulier : la partie intracellulaire des TM3 et 7.



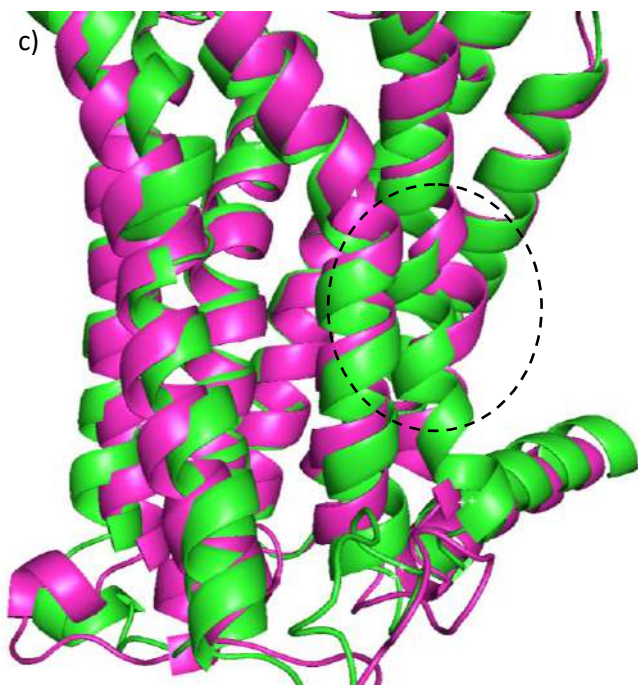
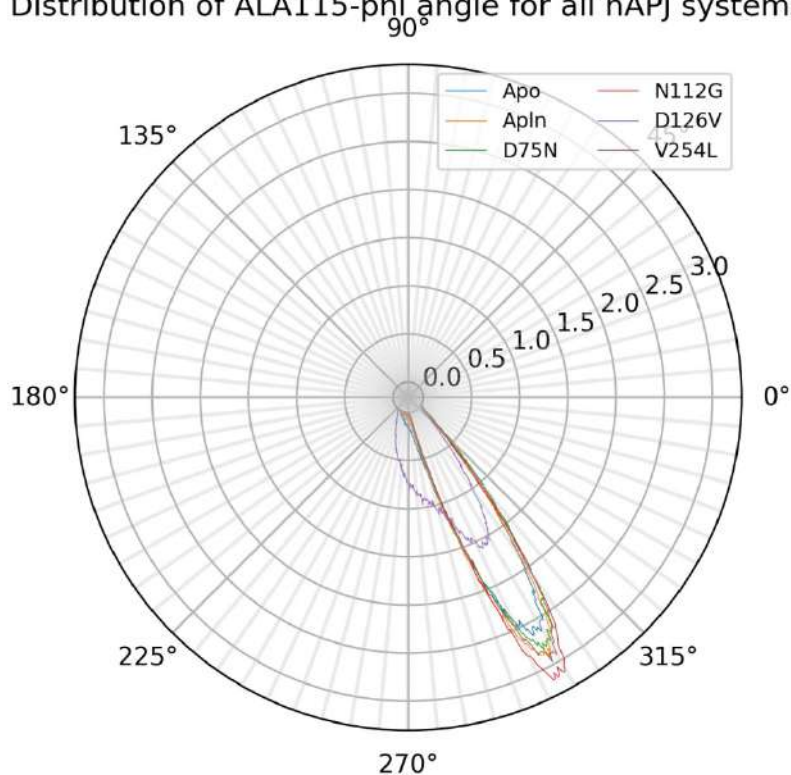


Figure 36 : Illustration des changements conformationnels distinguant le système hAPJ-D126V-apo du système hAPJ-WT-apo

a) Un des changements les plus significatifs (entouré) est l'éloignement relatif des TM3 et 6 dans le système hAPJ-D126V-apo (magenta, état 12) par rapport au système hAPJ-WT-apo (vert, état 11). L'explication de ce changement conformationnel se trouve dans l'environnement du résidu V126 lorsqu'il se trouve dans cette conformation. b) Dans cette conformation, le résidu V126 est entouré par des groupements hydrophobes : la chaîne latérale de F67, L142 et A149, ainsi que le squelette du résidu G146. c) L'autre changement le plus significatif (entouré, même code de couleurs et états) est celui de la partie intracellulaire du TM7. Ce changement est l'inverse de celui observé dans les systèmes hAPJ-N112G-apo et hAPJ-WT-apeline : au lieu d'une déformation vers le TM2 comme le mutant N112G, on observe une déformation dans le sens opposé, tel qu'observé à la figure 35 b) et d).

a) Distribution of ALA115- ϕ angle for all hAPJ systems



b) Distribution of SER116- ψ angle for all hAPJ systems

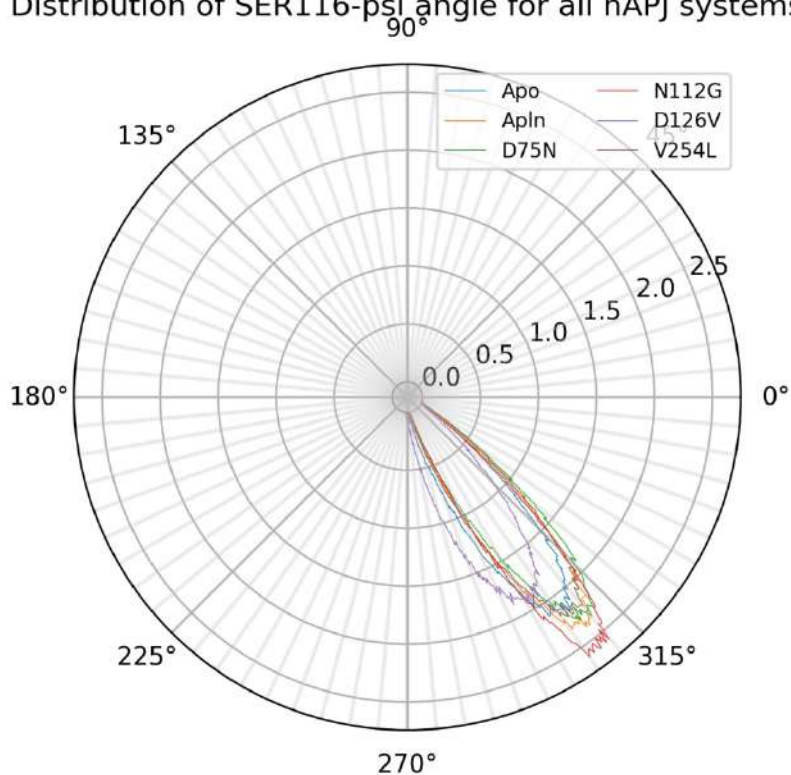
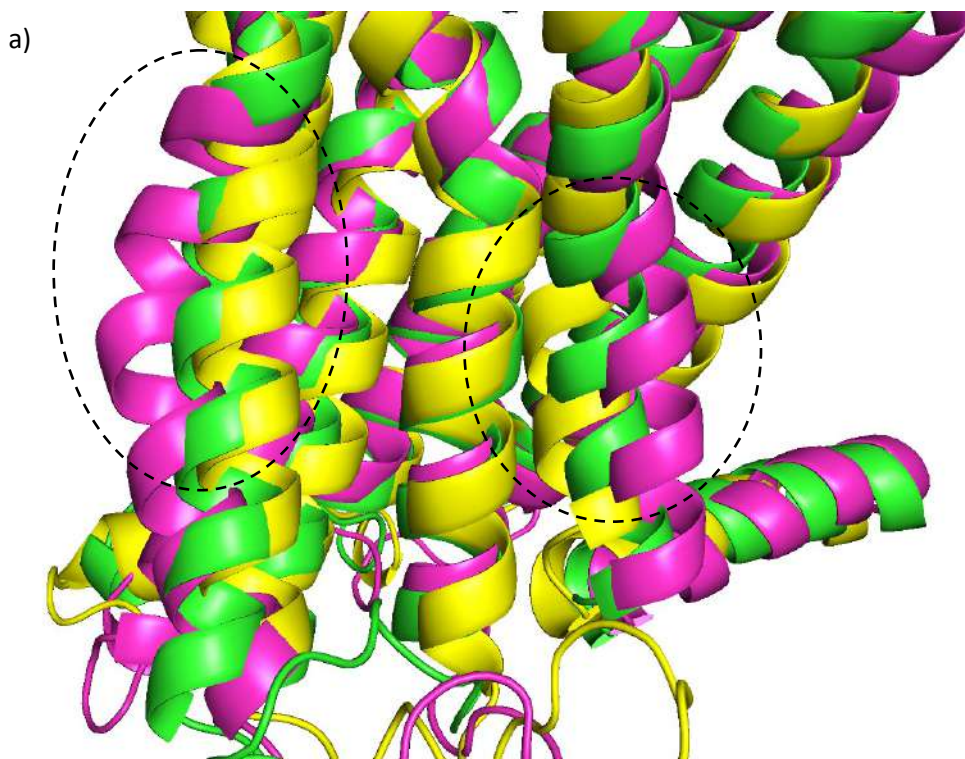


Figure 37 : Distribution des angles dièdres expliquant le changement conformationnel du TM3 pour le système hAPJ-D126V-apo

a) L'angle dièdre A115- ϕ varie considérablement par rapport aux autres systèmes. b) Il en va de même, de façon plus subtile, pour l'angle dièdre S116- ψ . Comme ces résidus se trouvent au milieu du TM3, leurs effets sur la position du résidu V126, situé en bas du TM6, sont importants.

Comparaison des systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ-V254L-apo

Les systèmes hAPJ-WT-apo et hAPJ-V254L-apo diffèrent principalement en deux endroits en particulier : le TM5 et la portion centrale du TM7. Au niveau du TM7, les angles dièdres semblent indiquer un mouvement similaire à celui des systèmes hAPJ-WT-apeline et hAPJ-N112G-apo, mais en réalité, ce mouvement dans le système hAPJ-V254L-apo est perpendiculaire à celui des deux autres.



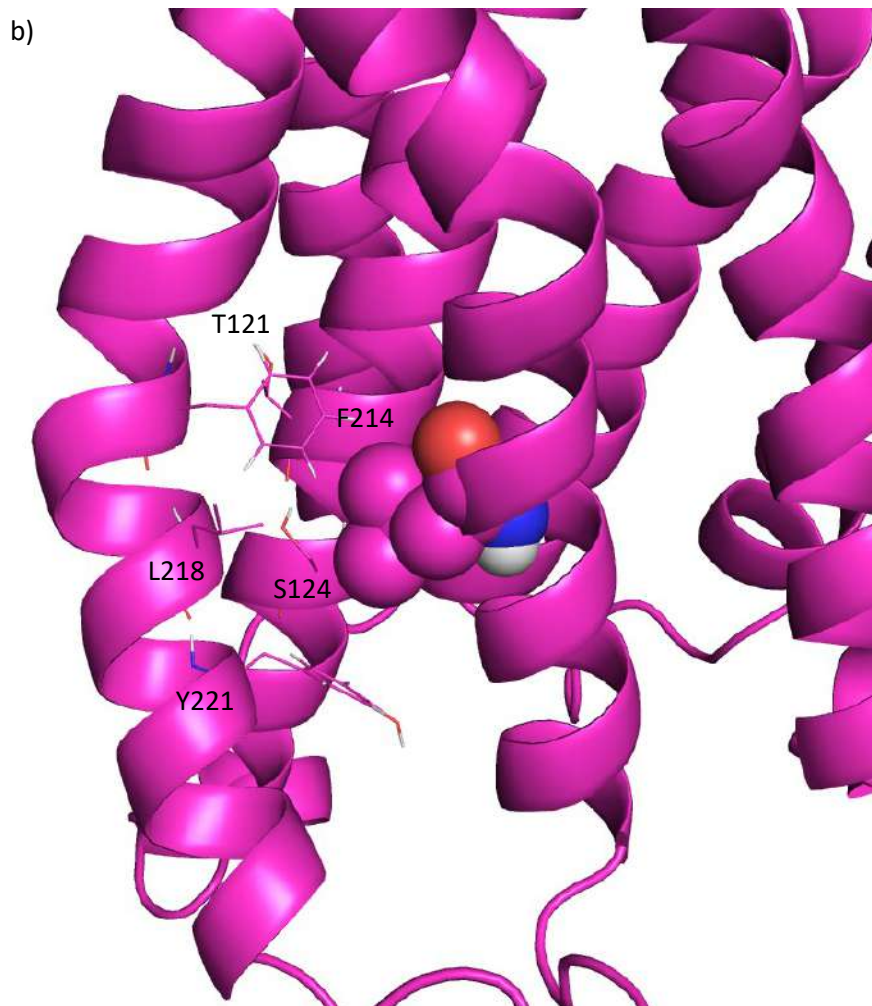
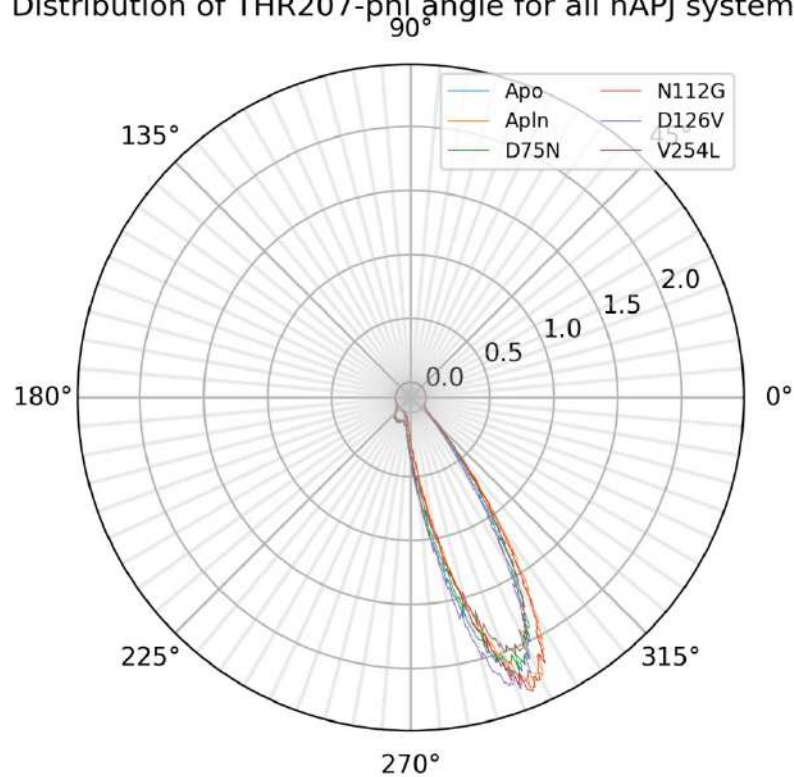


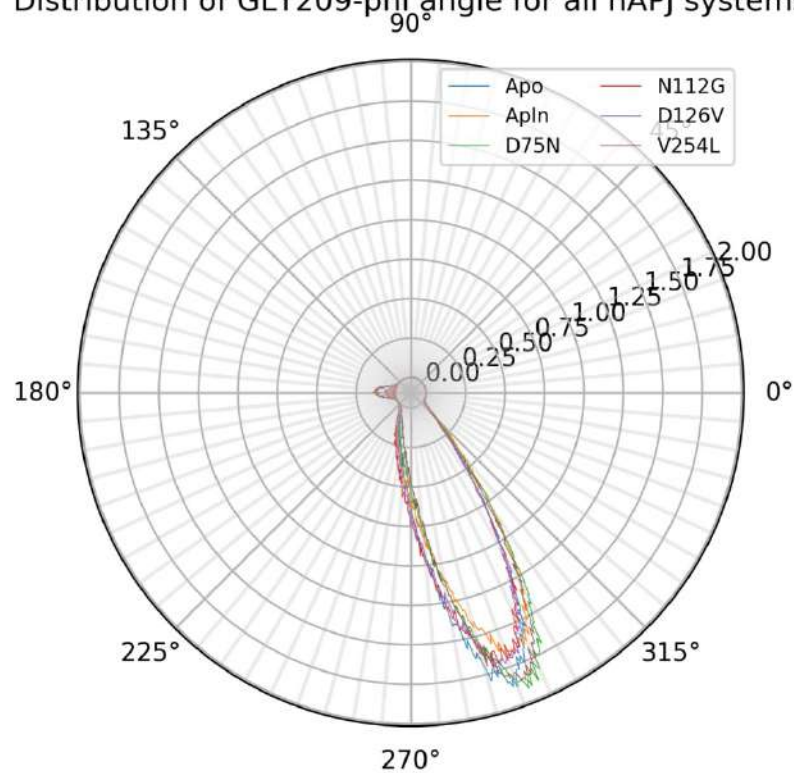
Figure 38 : Comparaison et explication des différences entre les systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ-V254L-apo

a) Comparaison des conformations des états hAPJ-WT-apeline (jaune, état 2), hAPJ-WT-apo (vert, état 11) et hAPJ-V254L-apo (magenta, état 5). La déformation du TM5 (encadré à gauche) est visible pour le système hAPJ-V254L-apo, mais pas pour les autres. La déformation du TM7 (encadré à droite) est différente pour les systèmes hAPJ-WT-apeline et hAPJ-V254L-apo. Dans le premier cas, la déformation rapproche le TM7 du TM2, tandis que dans le deuxième, la déformation le rapproche du TM1. b) Explication de la déformation du TM5 par la mutation V254L. Cette mutation rajoute du volume dans un espace déjà restreint. Les résidus T121, S124, F214, L218 et Y221 se trouvent à proximité et l'ajout de ce volume les pousse plus loin du TM6, créant ainsi une déformation.

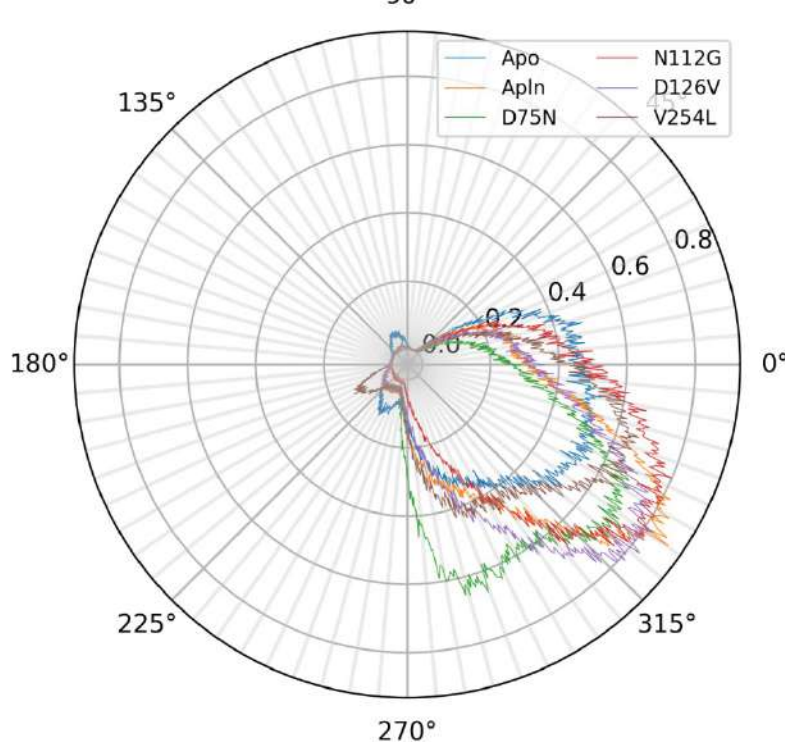
a) Distribution of THR207- ϕ angle for all hAPJ systems



b) Distribution of GLY209- ϕ angle for all hAPJ systems



c) Distribution of GLY209- ψ angle for all hAPJ systems



d) Distribution of PHE210- ϕ angle for all hAPJ systems

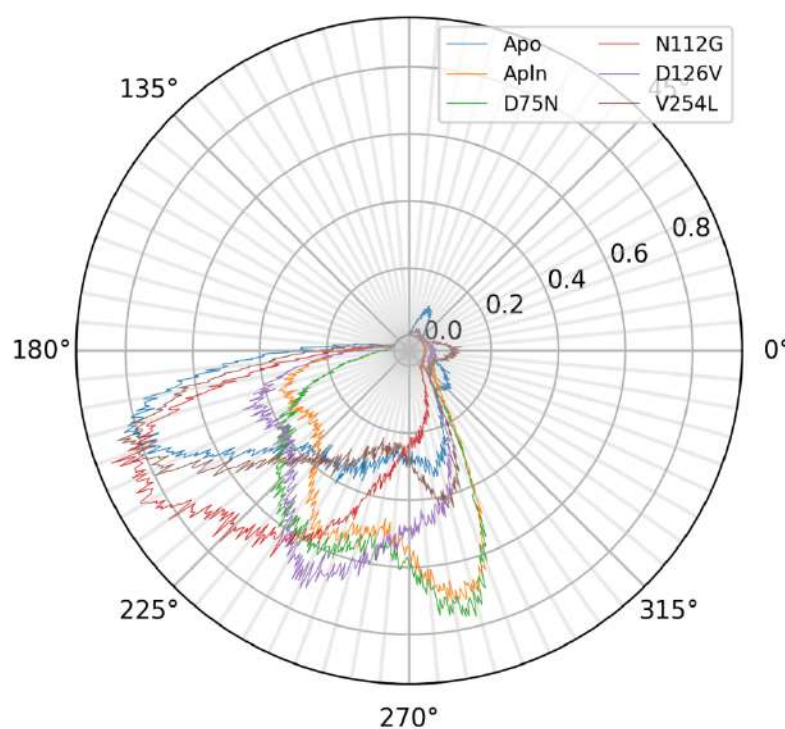


Figure 39 : Distribution des angles dièdres expliquant le changement conformationnel du TM3 pour le système hAPJ-V254L-apo

a) L'angle dièdre T207- ϕ présente une distribution bimodale uniquement dans le système V254L, indiquant la présence d'un changement conformationnel unique à ce système. b) L'angle dièdre G209- ϕ présente également une distribution bimodale dans le système V254L. Le mode le moins peuplé est plus important dans ce système que les autres, ce qui indique une certaine importance pour ce système. Les différences les plus importantes se manifestent cependant dans les distributions des angles dièdres c) G209- ψ et d) F210- ϕ . Ici, pour chaque distribution, le système V254L présente un mode différent des autres systèmes et dont la population est non-négligeable.

Discussion

Résumé des résultats

D'abord, l'analyse des simulations des systèmes hAPJ-WT-apo et hAPJ-WT-apeline ont permis d'identifier les régions du récepteur où s'effectuent les changements conformationnels les plus lents associés à son activation. Hormis les changements d'orientation et position relatives des TMs, on peut constater que les changements conformationnels sont concentrés au niveau du HHM et du MHBN. Ces régions et leur environnement ont donc été ciblés par une panoplie de mutations, dont D75N, N112G, D126V et V254L. Ces mutations ont exhibé un comportement déviant substantiellement des comportements des systèmes hAPJ-WT-apo et hAPJ-WT-apeline. D75N est une mutation favorisant les formes inactives du récepteur, tandis que N112G favorise plutôt ses formes actives. Les mutations D126V et V254L favorisent spécialement les formes actives du récepteur sont la signalisation s'effectue préférentiellement via les protéines G. Ces mutants ont ensuite été testés en dynamique moléculaire. Pour les systèmes hAPJ-D126V-apo et hAPJ-V254L-apo, il est observé que les changements conformationnels les plus importants sont concentrés au niveau de l'interface entre les TMs 3, 5 et 6. Pour le système hAPJ-N112G-apo, les changements conformationnels sont situés aux mêmes endroits que le système hAPJ-WT-apeline. Finalement, pour le système hAPJ-D75N-apo, les changements conformationnels sont différents de tous les autres systèmes et concentrés au niveau du TM4, du haut du TM5 et de l'hélice 8.

Critique du protocole d'échantillonnage

Un protocole d'échantillonnage de dynamique moléculaire bien conçu doit présenter les caractéristiques suivantes :

- Une réaction cohérente des propriétés des systèmes à une variation importante de leur environnement.
- Une approximation aussi juste que possible de la distribution à l'équilibre des systèmes.
- Une approximation aussi juste que possible de la dynamique conformationnelle du système.

Sans informations expérimentales, il est impossible de confirmer si le système théorique correspond au système expérimental, et donc pratiquement impossible de confirmer que la cinétique et la thermodynamique des états du système théorique est à l'équilibre. Cependant, il est possible de vérifier si le système réagit à des variations environnementales.

Si le protocole d'échantillonnage est bien conçu, les différences environnementales (i.e. mutations, ligands) auront un impact sur la dynamique et les ensembles conformationnels résultants. En observant les différences théoriques des conformations des différents systèmes, on constate que, bien que les systèmes partagent des ensembles conformationnels similaires, il existe des différences bien réelles dans leur dynamique et les ensembles conformationnels qu'ils explorent.

La comparaison des systèmes hAPJ-WT-apo et hAPJ-WT-apeline sur les aspects de l'analyse par sous-matrices principales et des ensembles conformationnels révèle des différences significatives quant aux résidus impliqués dans les changements conformationnels, ainsi que dans les changements conformationnels caractérisant le système, comme il est possible de voir dans les figures 15, 18, 19 et 20. De même, les systèmes hAPJ-N112G-apo, hAPJ-D126V-apo et hAPJ-V254L-apo sont issus d'un même état du système hAPJ-WT-apo. En observant les figures 27, 29, 31 et 34 à 39, on constate que, bien qu'issus d'un seul état du système hAPJ-WT-apo, chacun des systèmes a divergé significativement de sa conformation de départ, peuplant ainsi des états non seulement variés, mais distincts du système hAPJ-WT-apo.

Cela signifie que le protocole d'échantillonnage est capable de mettre en évidence l'effet de l'environnement sur un système. Sur cet aspect de la question, on peut donc conclure que le protocole est approprié. Cependant, même si les deux autres aspects de la qualité du protocole sont impossibles à évaluer, il existe des mesures qui pourraient être prises qui amélioreraient le protocole d'échantillonnage. Ces deux mesures sont l'utilisation d'un MSM en phase d'échantillonnage, pour ensuite retourner aux HMMs en phase d'analyse,

ou encore un meilleur choix de conformations représentatives à l'aide du HMM, ainsi que l'exploration de la relation entre différents résidus par la variation du délai τ .

Critique de la tICA comme outil d'analyse des données

Bien que la tICA corresponde essentiellement au type d'analyse recherchée dans ce travail, elle est, comme toute forme d'analyse, imparfaite. Un premier aspect important à considérer lors de l'utilisation de la tICA est que, lorsque le temps vient de comparer différents systèmes, l'analyse ne considère que les variations par rapport à la moyenne. Autrement dit, les résultats de la tICA sont invariants sous les translations. Cela signifie que toute comparaison de projection de séries de données sur les composantes de leur tICA respective empêche toute forme de comparaison de moyenne : la moyenne des deux séries de données après leur projection se trouve à l'origine.

Un inconvénient potentiel de la technique est sa dépendance à la linéarité des changements conformationnels fortement autocorrélés. En effet, l'équation qui définit la matrice d'autocorrélation est en essence la même que l'équation de corrélation linéaire entre deux variables. Elle présente donc les mêmes faiblesses. Par exemple, considérons le processus suivant :

$$\left\{ \underbrace{(1,0), \dots, (1,0)}_{N \text{ fois}}, \underbrace{(0,1), \dots, (0,1)}_{N \text{ fois}}, \underbrace{(-1,0), \dots, (-1,0)}_{N \text{ fois}}, \underbrace{(0,-1), \dots, (0,-1)}_{N \text{ fois}}, \underbrace{(1,0), \dots, (1,0)}_{N \text{ fois}}, \dots \right\}$$

La partie de la matrice d'autocorrélation dans le sens de l'écoulement du temps avec un délai $\tau = N + 1$ vaut :

$$\begin{aligned} & \frac{1}{4N} \left(N \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} + N \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} -1 & 0 \end{pmatrix} + N \begin{pmatrix} -1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \end{pmatrix} + N \begin{pmatrix} 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} \right) \\ & = \frac{1}{4} \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \end{aligned}$$

Cette matrice est antisymétrique et, par conséquent, la somme des matrices d'autocorrélation dans les deux sens de l'écoulement du temps donne la matrice nulle :

$$C_s(N+1) = \frac{1}{4} \left(\begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix} \right) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Il est pourtant clair que ce processus présente des états fortement corrélés, mais cette corrélation n'est pas linéaire, elle est périodique. La tICA telle qu'utilisée n'identifie que les autocorrélations à un temps précis, ce qui signifie qu'un mouvement périodique peut avoir une autocorrélation faible ou élevée dépendamment du τ choisi. Ceci dit, à l'échelle de temps utilisée, la probabilité qu'il existe un mouvement à période stable dans le système est pratiquement nulle. Il n'en est pas moins important de garder cette information en tête lorsque le temps de décalage est réduit, ou lorsqu'on considère des changements conformationnels impliquant plusieurs minima énergétiques locaux le long d'une coordonnée de réaction non-linéaire. De plus, dans le thème de la période, l'utilisation de la tICA masque la véritable nature de la technique. Soit la trajectoire $\mathbf{x}(t)$ fonction matricielle $M^*(\tau)$, avec \star l'opérateur d'autocorrélation :

$$M^*(\tau) = \frac{1}{2} \left[(x_i \star x_j)(\tau) + (x_j \star x_i)(\tau) \right]_{n \times n}$$

$$(x_i \star x_j)(\tau) = \mathcal{F}^{-1} \left(\overline{\mathcal{F}(x_i)} \mathcal{F}(x_j) \right) (\tau)$$

\mathcal{F} représente ici la transformée de Fourier (discrète ou continue), \bar{x} la conjuguée complexe de x et \mathcal{F}^{-1} la transformée de Fourier inverse. La matrice analysée par la tICA n'est en réalité qu'une infime partie de $M^*(\tau)$, avec le τ d'intérêt fixé (ici, par exemple, autour de 100 ns). Cela signifie qu'une part importante des relations entre les données est masquée par la technique. Par exemple, si on suppose que la dynamique conformationnelle d'un certain résidu affecte celle d'un autre résidu, et ainsi de suite jusqu'à un résidu important d'un site actif, cette cascade serait visible dans $M^*(\tau)$ avec un τ faible, et à mesure que τ augmente, sous la forme d'une augmentation de l'autocorrélation entre le résidu important et le premier résidu de la cascade, puis le deuxième, et ainsi de suite jusqu'à l'avant

dernier. Ce type d'influences entre résidus peut être observé avec la tICA, mais plus spécifiquement, l'analyse cherchant à cibler ce type d'influences est l'analyse cyclostationnaire (Ho et Hamelberg, 2016), qui utilise une forme limitée de $M^*(\tau)$. En effet, le problème avec $M^*(\tau)$ est que son temps de calcul est excessivement long : il est de l'ordre de $O(n^2 N \log N)$, avec n le nombre de paramètres conformationnels (e.g. angles dièdres) et N le nombre d'images de simulation accumulées.

L'analyse cyclostationnaire, ou d'autres analyses du même genre impliquant l'autocorrélation, pourraient permettre d'identifier des réseaux de résidus dont les changements conformationnels sont corrélés au-travers d'un GPCR sans assumer une distribution stationnaire des données, ce qui complète en quelque sorte la tICA. Ce type d'analyse pourrait être effectué à l'avenir afin d'étudier plus en détails la relation entre différents résidus de hAPJ. Un aspect important à considérer si ce type d'expérience vient à être mené est le taux d'échantillonnage des images de dynamique moléculaire. En effet, le théorème d'échantillonnage de Shannon-Nyquist stipule clairement que pour qu'un signal de fréquence f soit bien défini, $2f$ est la borne inférieure stricte de l'échantillonnage. L'identification de corrélations entre résidus s'arrête donc à un τ égal à deux fois le temps entre deux images de la simulation. Une alternative à ce type d'analyse, plus robuste, a récemment été publiée et pourrait également être utilisée (Doddi *et al*, 2019).

Critique de l'analyse par sous-matrices principales de la tICA

Pour évaluer l'analyse proposée, les représentations des résultats de l'analyse de la tICA ont été comparées avec les distributions d'angles dièdres et les changements conformationnels observés chez les conformations représentatives de chaque système (Figure 18 et 27 vs 34 et 35, 25 vs 32 et 33, 29 vs 36 et 37, 31 vs 38 et 39). On peut constater, en particulier pour les pics détectés dans la portion centrale du récepteur, soit là où les efforts de recherche ont été concentrés, qu'un changement conformationnel est observable à au plus un tour d'hélice autour de chacun des pics. Ainsi, la technique d'analyse par sous-matrices principales peut permettre de mieux repérer des changements conformationnels pertinents du récepteur, car le système reste hautement dynamique et les différentes conformations représentatives du système présentent de nombreux changements

conformationnels non-pertinents issus de composantes indépendantes de plus basse autocorrélation qui ne sont pas considérées lors de l'optimisation du système. Cependant, comme la technique a seulement été utilisée sur les différentes variantes du récepteur APJ, dont on ne connaît que très peu la dynamique des conformations, l'analyse par sous-matrices principales doit encore être utilisée sur d'autres systèmes dont on connaît mieux le comportement avant d'être complètement validée. Par exemple, la simulation de 1 ms de la protéine BPT1 pourrait être analysée (Shaw *et al*, 2009), ou encore d'autres simulations décrites dans McGibbon *et al*, 2014.

Critique des HMM en phase d'échantillonnage

Comme il a été expliqué dans l'introduction, les GMHMMs utilisent un mélange de Gaussiennes pour décrire les états du système. Les Gaussiennes ont un seul mode, qui correspond à la moyenne. En théorie, l'image de simulation représentative de cet état se trouve à la moyenne. En pratique, c'est l'image la plus proche qui est choisie, ce qui permet d'évaluer la diversité des conformations du système en examinant les conformations les plus probables pour chaque état. Si cette estimation est particulièrement appropriée pour l'analyse des systèmes, elle l'est cependant beaucoup moins dans un contexte d'échantillonnage.

Le problème majeur avec l'utilisation d'un GMHMM pour l'échantillonnage est précisément ce qui le rend si pratique pour l'analyse d'un système : les images sélectionnées se rapprochent du mode de la distribution associée à l'état. Or, cela signifie que ces images sont les plus probables, donc les plus stables, de cet état. Cela contrevient au principe de sélection des états les moins échantillonnés. Ces états les moins échantillonnés sont moins probables, donc plus instables. Cependant, ce principe est exploité de façon optimale lorsque les états sélectionnés sont non seulement les moins échantillonnés, mais également les plus loin des maxima locaux de la distribution des images du système. Cela s'explique par le fait que, pour tout changement conformationnel ou réaction chimique, l'état le plus instable le long de la coordonnée de réaction est l'état de transition. Par conséquent, le principe de choix des états les moins échantillonnés se base sur l'hypothèse que ces états sont plus près de maxima énergétique locaux, donc de

potentiels états de transition. Un échantillonnage plus approfondi autour des images provenant de ces états a donc plus de chances d'échantillonner un état du système jusque-là non-découvert par la simulation.

De par leur nature, les HMMs identifient des états dont la moyenne est un maximum local de la distribution à l'équilibre du système. Pour identifier les états les moins échantillonnés, il est nécessaire de soit bonifier l'analyse telle quelle, soit ajouter des informations. Ici est incluse une suggestion qui se base sur le HMM pour sélectionner quoi échantillonner.

Cette suggestion utilise la distance de Mahalanobis, une version multidimensionnelle de la cote Z :

$$d_{Mij} = \sqrt{(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)}$$

La suggestion fonctionne ainsi, en utilisant comme coordonnées la projection des données sur les composantes de la tICA au lieu du système entier :

- Classer les images de simulation en ordre décroissant en fonction de leur distance minimale d'un état du HMM : $d_{Mi} = \min_j d_{Mij}^2$
- Choisir l'image dont la distance minimale d'un état du HMM est maximale : $x_m = \max_i (d_{Mi})$. Cette image est soit une conformation hautement énergétique ou dans un état dont la population est suffisamment faible pour avoir été ignorée par le protocole.
- Classer les images de simulation en ordre décroissant en fonction de leur distance de cette image $d_{im} = (x_i - x_m)^T (x_i - x_m)$
- L'image dont la somme des rangs des deux distances (d_{Mi} et d_{im}) est la plus faible (rang 1 étant le maximum de chaque distance) devient la seconde image à échantillonner.
- Ainsi de suite, on identifie les états à échantillonner en additionnant les rangs de la distance de Mahalanobis et des distances entre chaque point et les images précédemment sélectionnées. Au besoin, le rang de la distance de Mahalanobis peut

être multiplié par le nombre de points sélectionnés, de façon à éviter la diminution de l'importance de ce critère.

Cet algorithme s'assure d'abord que les images sélectionnées se trouvent autant que possible hors de la région de confiance de chacune des Gaussiennes qui définissent les différents états. Ensuite, il s'assure que chaque nouvelle image sélectionnée se trouve aussi loin que possible des images précédemment sélectionnées. Ainsi, l'algorithme choisit les images les plus diverses et les plus instables possible en fonction des analyses disponibles, ce qui correspond bien au principe d'échantillonnage des états les moins échantillonnés. Une autre solution est d'utiliser les MSMs, qui contiennent naturellement des états très peu échantillonnés et hautement énergétiques et se prêtent ainsi mieux à la sélection des états les moins échantillonnés.

Critique de l'utilisation de la distance de Bhattacharyya pour séparer les états

Un défaut majeur de l'approche utilisée pour séparer les états n'est pas la distance de Bhattacharyya elle-même, mais comment elle a été utilisée. En effet, un de ces principaux attraits est qu'il est possible d'isoler la contribution de chaque distance des autres composantes de cette distance. Cet attrait n'a pas été utilisé à son plein potentiel et explique la déconnexion entre l'analyse graphique de la distribution des images des simulations, l'analyse des conformations représentatives de ces états et l'analyse des diagrammes d'angles dièdres. En effet, pour avoir un intérêt fonctionnel, une conformation doit être stable. La population d'une conformation transitoire, qui existe de façon stable pendant un intervalle de temps extrêmement court avant un nouveau changement conformationnel d'envergure, est trop faible pour s'associer efficacement à d'autres protéines et exécuter sa fonction. Il faut au moins que la conformation ait une population significative pour être intéressante et cela signifie que cette conformation ne doit pas faire partie d'un ensemble conformationnel à interconversion rapide qui ne partage pas les mêmes propriétés, sans quoi ses particularités ne deviennent guère plus intéressantes qu'un artefact. C'est là que l'analyse graphique des états laisse paraître toute sa faiblesse : le critère de sélection des états stables est totalement arbitraire. Comme les gaussiennes utilisées pour représenter les puits énergétiques du système ont une matrice des covariances

diagonale, l'analyse de la forme des puits énergétiques qu'elles identifient a une portée très limitée. De plus, il n'est pas garanti qu'un puits éloigné d'un autre n'échange pas fréquemment et directement des conformations. Il n'est donc pas approprié d'utiliser une méthode de séparation des états qui dépend de l'ampleur des changements conformationnels et de la forme des distributions utilisées. Ceci dit, plusieurs solutions s'offrent à nous. La première utilise le temps de mélange (*mixing time*) du HMM (Levin *et al*, 2008) :

$$t_{mix} = \min_{t \in \mathbb{N}} \left(\frac{1}{2} (\mathcal{P}^t(\tau) \mathbf{s}) \cdot \boldsymbol{\mu} \leq \frac{1}{4} \right)$$

En choisissant $\mathbf{s}^T = (0, \dots, 1, \dots, 0)$, soit la distribution où l'état d'intérêt est le seul état occupé, t_{mix} devient le temps nécessaire à cet ensemble conformationnel pour se mêler au reste des états, en général.

Une autre option, plus spécifique, consiste à choisir la demi-vie d'un état (équ. 9) :

$$\begin{aligned} P(Y_{t+\tau} = S_i | Y_t = S_i) &= a = \mathcal{P}_{ii}(\tau) \Rightarrow P(Y_{t+\tau} \neq S_i | Y_t = S_i) = 1 - a \\ \Rightarrow P\left(Y_{t+\frac{t_1}{2}} \neq S_i \middle| Y_t = S_i\right) &= (1 - a)^{\frac{t_1}{2}} = \exp\left(\frac{t_1}{2} \lambda_i\right) = \frac{1}{2} \\ \Rightarrow \lambda_i &= \frac{\ln(1 - a)}{\tau}, t_{\frac{1}{2}} = -\frac{\ln(2)}{\ln(1 - a)} \end{aligned}$$

Pour un ensemble d'états, il est difficile d'identifier la probabilité de sortie de l'ensemble, puisque cette probabilité dépend des populations de l'ensemble évalué. Une façon d'évaluer la demi-vie d'un ensemble est de présumer que cet ensemble est à l'équilibre dans le contexte de la chaîne de Markov plus large et stationnaire tout le long du processus :

$$\boldsymbol{\mu}_S = \begin{cases} \frac{\mu_i}{\sum_{s_i \in S} \mu_i}, s_i \in S \\ 0, \text{sinon} \end{cases}$$

$$P(Y_{t+\tau} \in S | Y_t \in S) = \sum_{s_i \in S} \mathcal{P}_i(\tau) \mu_s = a_s$$

$$\Rightarrow \lambda_s = \frac{\ln(1 - a_s)}{\tau}, t_{\frac{1}{2}} = -\frac{\ln(2)}{\ln(1 - a_s)}$$

$\mathcal{P}_i(\tau)$ représente ici la i -ième ligne du propagateur.

Si ces méthodes ne sont pas suffisantes ou se révèlent être inappropriées, c'est à ce moment qu'une analyse graphique peut être utilisée pour confirmer le résultat de la mesure. Pour ce faire, l'analyse graphique doit également être améliorée. Il doit être possible d'observer la séparation d'un état d'intérêt et des autres états, en termes de distance de Bhattacharyya, sans simplement se concentrer sur les graphiques de paires de dimensions adjacentes; il faut identifier l'ensemble minimal de composantes indépendantes permettant d'isoler cet état et effectuer l'analyse sur cet ensemble. Ce problème est identique au problème de couverture par ensembles et est NP-complet, mais un algorithme glouton donne une bonne approximation de la solution optimale : en commençant par la composante qui sépare le plus d'états de l'état d'intérêt, on sélectionne ensuite la composante qui sépare le plus d'états encore reliés à celui d'intérêt, et ainsi de suite jusqu'à ce qu'il ne reste que l'état d'intérêt. Bien sûr, si l'état d'intérêt est indiscernable d'un autre état, cet algorithme peut être appliqué à l'ensemble des états confondus au lieu d'un seul état pour des résultats similaires. Cependant, les deux états confondus doivent être considérés conjointement.

Critique des paramètres structuraux utilisés

Un des désavantages majeurs de l'utilisation des angles dièdres sans autre paramètre structural est qu'il devient difficile d'observer des changements conformationnels des structures secondaires plus longues. En effet, les résidus qui délimitent et constituent ces structures sont rarement rigides et leurs angles dièdres varient de façon significative. Pour percevoir un changement conformationnel, la moyenne d'un angle dièdre doit varier de façon significative, mais dans le cas de longues structures secondaires, un changement conformationnel important peut être la conséquence de petites variations d'angles dièdres tout le long de la structure. Ce type de changements conformationnels peut ne pas être détecté efficacement par l'analyse seule des angles dièdres. Dans le cas particulier des

GPCRs, une solution potentielle est de considérer chaque domaine transmembranaire de façon séparée. En définissant les limites de chaque domaine, il est possible de dégager les axes principaux de la structure en effectuant par exemple une PCA sur les positions des carbones α . L'axe principal \mathbf{a} est celui avec la plus grande variance, et la variance des axes secondaires donnent le degré de distorsion D du domaine transmembranaire :

$$\Sigma_{TM} = \sum_{i \in TM} (\mathbf{x}_i - \boldsymbol{\mu}_{TM})(\mathbf{x}_i - \boldsymbol{\mu}_{TM})^T = V\Lambda V^T = (\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3) \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \mathbf{v}_3^T \end{pmatrix}$$

$$\mathbf{v}_1 = \mathbf{a} \quad D = \sqrt{\lambda_2 + \lambda_3}, \quad \lambda_1 > \lambda_2 > \lambda_3$$

Un autre désavantage important des angles dièdres, est qu'il leur est difficile de détecter les changements de structure tertiaire. De la même façon que les structures secondaires, les structures tertiaires peuvent être modifiées par plusieurs petites variations d'angles dièdres. Cette fois, la façon de détecter ces mouvements dans le cas des GPCRs est d'analyser la distance et l'élévation relatives des domaines transmembranaires. Une façon d'obtenir ces mesures commence par le calcul du centre de masse des carbones α du récepteur, tout comme celui de chaque domaine transmembranaire. Les axes principaux du GPCR peuvent ensuite être calculés de la même façon que pour chaque TM.

L'élévation relative h_{ij} et la distance relative r_{ij} des domaines transmembranaires peuvent ensuite être calculées :

$$\Sigma_{rec} = \sum_{i \in rec} (\mathbf{x}_i - \boldsymbol{\mu}_{rec})(\mathbf{x}_i - \boldsymbol{\mu}_{rec})^T = U\Lambda U^T = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3) \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{pmatrix}$$

$$\lambda_1 > \lambda_2 > \lambda_3$$

$$d_{ij} = \left| \text{proj}_{\{\mathbf{u}_2, \mathbf{u}_3\}} (\boldsymbol{\mu}_{TM_i} - \boldsymbol{\mu}_{TM_j}) \right|$$

$$h_{ij} = \left| \text{proj}_{\mathbf{u}_1} (\boldsymbol{\mu}_{TM_i} - \boldsymbol{\mu}_{TM_j}) \right|$$

L'opérateur $\text{proj}_{\{\mathbf{v}_2, \mathbf{v}_3\}}$ signifie ici la projection sur le plan défini par les vecteurs \mathbf{v}_2 et \mathbf{v}_3 . Le calcul des axes principaux du GPCR permettent d'éliminer l'impact de l'orientation de

la protéine dans la membrane sur les coordonnées brutes. Suivant ce calcul, chaque domaine transmembranaire peut être assigné une inclinaison par rapport aux axes principaux du GPCR :

$$U^T \mathbf{a} = \mathbf{b}$$

$$\mathbf{b} = \begin{pmatrix} \sin(\varphi_a) \cos(\theta_a) \\ \sin(\varphi_a) \sin(\theta_a) \\ \cos(\varphi_a) \end{pmatrix}$$

En assumant que $\sin(\varphi_a) \geq 0$, ce qui est raisonnable puisque \mathbf{a} et $-\mathbf{a}$ font autant office d'axes principaux, la résolution de ces équations pour trouver φ_a et θ_a est triviale. Le dernier degré de liberté rotationnel (rotation selon l'axe principal \mathbf{a}) est plus complexe et requiert une référence interne à chaque domaine. Par exemple, dans le cas des GPCRs, cela pourrait être le résidu n.50 de chaque TM. L'orientation relative des résidus peut alors être calculée, toujours en prenant comme référence les axes principaux du récepteur :

$$\mathbf{c}_n = \text{proj}_{\{u_2, u_3\}}(\boldsymbol{\alpha}_{n.50} - \boldsymbol{\mu}_{TMn})$$

$$\cos(\omega_{ij}) = \frac{\mathbf{c}_i \cdot \mathbf{c}_j}{|\mathbf{c}_i| |\mathbf{c}_j|}$$

L'orientation relative de chaque paire de TMs peut être calculée ainsi. L'ensemble de ces mesures devrait donner une bonne idée de la conformation globale d'un GPCR : les distances relatives, les inclinaisons et les orientations relatives, ainsi que le niveau de distorsion de chaque domaine transmembranaire sont décrites par ces mesures, ce qui ajoute un total de 70 dimensions supplémentaires. Considérant que la tICA n'est pas une étape limitante de l'analyse, la perte de performance est minime sur ce point. Cependant, le calcul des axes principaux du GPCR pour chaque image de la simulation peut être prohibitif : le temps de calcul est de l'ordre de $O(nN)$, avec n le nombre de résidus du récepteur et N le nombre d'images.

Analyse des résultats expérimentaux

Critique de la qualité et limitations des résultats expérimentaux

Les résultats expérimentaux obtenus dans le cadre de ce travail présentent plusieurs limitations qui restreignent leur interprétation à un contexte spécifique.

La première limitation se trouve dans la nature des systèmes étudiés. Il s'agit d'expériences effectuées sur une seule lignée clonale (HEK293), donc dans un environnement cellulaire différent de celui de n'importe quelle cellule humaine et dans un contexte réductionniste où les protéines étudiées sont surexprimées par rapport à toutes les autres. Or, même si les protéines d'intérêt sont surexprimées, l'environnement cellulaire a un impact sur leur comportement. En effet, les GPCRs n'interagissent pas qu'avec les protéines G et les arrestines. Tel que mentionné dans l'introduction, toute une gamme de protéines participe à leur régulation et leur transit dans la cellule, incluant par exemple les GRK, RGS et les protéines des *clathrin-coated pits*. La réaction des GPCRs à la présence de ligands dans le système étudié n'est donc aucunement représentative de la véritable chronologie ou de la véritable amplitude de la réaction avec un niveau d'expression normal de toutes les protéines impliquées et effectuée chez l'humain c'est là le biais systémique associé à ces expériences.

Un aspect particulièrement préoccupant de cette limitation se manifeste dans les essais BRET. Dans le cas de la β -arrestine-2, tel que mentionné dans l'introduction, les arrestines ont une affinité accrue pour les GPCRs phosphorylés. Dans un contexte de surexpression du GPCR et de la β -arrestine-2, mais pas de la GRK, il est possible que les observations soient faussées. En effet, les paramètres pharmacologiques d'essais BRET β -arrestine effectués en présence ou en absence de GRK peuvent donner des résultats différents, dépendant du récepteur et des ligands utilisés (Ibsen *et al*, 2019). Dans cette référence en particulier, c'est la cinétique qui est mesurée, mais des informations sur l'intensité de la fluorescence après 10 min d'incubation avec le ligand sont également présentes et indiquent qu'il existe bien une différence pour les essais BRET en état stationnaire tels qu'effectués ici. Cette source d'erreur peut également se propager aux essais $G\alpha_{i1}$: comme l'arrestine interrompt la signalisation via les protéines G, une chronologie anormale de

l'activation des arrestines peut également fausser les données pour les essais sur les protéines G. L'expression déficiente des RGS peut également poser problème dans le cadre de ces essais.

Une autre limitation des essais biologiques effectués sont la température et les conditions physico-chimiques dans lesquelles ils sont effectués. En effet, la température d'incubation et d'observation des essais de liaison et des essais BRET est la température de la pièce, soit entre 21 et 23°C. Cette température est loin des 37°C du corps humain, qui est aussi la température à laquelle les systèmes ont été simulés par dynamique moléculaire. Les ELISAs de surface étant effectués sur des cellules fixées directement sorties de l'incubateur à 37°C, leurs données ne sont pas soumises à cette source d'erreur. Cet écart de température peut avoir des effets significatifs sur les paramètres de liaison et de signalisation observés (Herenbrink *et al*, 2016). En effet, le système étudié est encore soumis aux lois de la physique, et chaque réaction, de même que chaque molécule, incluant le récepteur, le ligand et les membranes, a une constante cinétique qui change en fonction de la température, de façon plus ou moins conforme à l'équation d'Arrhénius :

$$k = A \cdot \exp\left(-\frac{E_a}{k_B T}\right)$$

La constante cinétique peut ici représenter toutes sortes de processus, comme la formation d'un radeau lipidique sur une membrane, ou encore les changements conformationnels d'un récepteur. Cela signifie que les différents équilibres qui régissent le comportement du système étudié sont modifiés par une différence de température de façon différente, dépendant des énergies d'activation (E_a) de chacune de ces réactions. Ainsi, il est possible que les résultats observés en BRET et les résultats observés en dynamique moléculaires ne correspondent pas. Cependant, il n'en reste pas moins que les conformations accessibles par un récepteur ne changent pas en fonction de la température. Ce sont plutôt les populations de chacune de ces conformations qui changent. À moins que l'effet du reste du système n'interfère de façon importante avec les changements conformationnels du récepteur (e.g. raidissement des membranes), il est peu probable que les résultats du BRET

soient non-représentatifs de ce qui s'observe en dynamique moléculaire. On peut conclure la même chose des essais de liaison : le processus de liaison s'effectue plus lentement à plus basse température, mais s'effectue quand même. Comme la constante de dissociation n'est pas calculée, mais plutôt le déplacement de l'apeline radiomarquée par l'apeline est observé, l'impact d'une variation de la constante de dissociation sur les résultats est minimisé.

Pour ce qui est des conditions physico-chimiques limitant l'interprétation des résultats, elles se divisent en deux parties : l'environnement biochimique et l'environnement inerte (i.e. ions, surfaces, etc.) L'environnement biochimique consiste en l'ensemble des protéines et molécules organiques présentes dans le milieu cellulaire, ce qui peut encore se diviser en deux parties : le liquide de culture et la surface d'adhésion. Que ce soit dans un milieu constitué de DMEM avec FBS, ou encore dans de la HBSS, la composition du liquide dans lequel baignent les cellules est différente de celle d'un humain, soit principalement, dans le cas qui nous intéresse, le sang et le liquide interstitiel.

Pour ce qui est de la surface d'adhésion, les puits de culture ont une surface soumise au traitement standard pour culture de tissus, soit du polystyrène traité avec un plasma contenant de l'oxygène sous basse pression (*vacuum-gas plasma*), ce qui permet d'oxyder la surface et de la rendre hydrophile. À cette surface adhèrent ensuite des protéines du milieu de culture, en particulier la vitronectine et la fibronectine, qui permettent finalement l'adhésion des cellules (Ryan, 2008). La quasi-totalité de la matrice extracellulaire, y compris les cellules qui en font partie (e.g. les fibroblastes), est par conséquent absente de l'environnement de culture. Considérant le rôle de la matrice extracellulaire dans une vaste gamme de processus biologiques (Hynes, 2009), son absence a nécessairement un impact sur le comportement des cellules étudiées.

Finalement, lorsqu'on parle d'environnement inerte, on parle principalement des ions présents en solution, et plus particulièrement du sodium. En effet, le sodium est connu pour moduler l'activité des GPCRs via un site de liaison allostérique situé autour du résidu 2.50 (D75 pour hAPJ) (Vsevolod *et al*, 2014; Liu *et al*, 2012; Schiffmann et Gimpl, 2018; White *et al*, 2018). Cet effet sera discuté plus en détail plus loin.

Ainsi, on constate que les mesures expérimentales obtenues au cours de ce travail ne peuvent être extrapolées vers quelque chose de plus général, soit un organisme vivant. Le nombre de facteurs ignorés est simplement trop grand. Même dans le contexte qui nous intéresse, celui de l'étude de l'interaction entre APJ et l'apeline-13, il est probable que les résultats soient influencés par un biais systémique ou observationnel. Cela signifie que, pour avoir le cœur net sur le véritable effet de ces mutations sur la signalisation d'APJ, que ce soit *in vitro* ou *in vivo*, il est nécessaire d'explorer davantage l'effet de ces mutants, que ce soit avec des concentrations variables de sodium, dans différents systèmes, si possible avec les GRKs, RGSs et autres protéines régulatrices, ou encore *in vivo* chez la souris.

Une autre et dernière source d'erreur dans ce travail est l'erreur humaine. En effet, une manipulation trop brusque des cellules et un temps trop long entre les sous-cultures ont possiblement nui à l'expression des protéines d'intérêt et à la multiplication des cellules, ce qui a possiblement affecté la qualité des résultats.

Analyse de la qualité et la validité des résultats des essais de liaison compétitifs

En observant les résultats des essais de liaison compétition, il est possible de constater que certains résultats de non-déplacement apparents sont dus à un rapport bruit/signal trop important, par exemple pour le mutant W261F. Au lieu d'observer le R^2 , il est possible d'utiliser le coefficient de corrélation des rangs de Spearman pour déterminer si la radioactivité suit une relation monotone décroissante avec la concentration du ligand. Les seuls mutants dont ce coefficient ne dépasse pas le seuil de significativité sont N112W, D75N et W261R. Les mutants seront classifiés et étudiés en fonction de si les données de radioactivité suivent précisément une relation michaélienne avec la concentration de ligand (liants), si elles suivent au moins une relation monotone décroissante, sans pouvoir conclure qu'elle soit michaélienne (« quasi-liants »), ou si elles n'ont aucune relation (non-liants).

Dans le cas des mutants quasi-liants et non-liants, plusieurs explications sont plausibles pour expliquer leur comportement :

1. Des erreurs expérimentales ont été effectuées et ont faussé les résultats,

2. Le récepteur mutant est instable sous les conditions de l'essai de liaison (e.g. dénaturation rapidement),
3. Le récepteur mutant est peu ou pas exprimé à la surface de la cellule,
4. La constante de dissociation de l'apeline radiomarquée a diminué pour ce mutant,
5. Le récepteur mutant est délétère pour la cellule, ou encore ralentit sa croissance.

L'instabilité induite par mutation s'appuie sur le principe que les formes dénaturées des protéines font partie de leurs conformations accessibles. Une déstabilisation uniforme des différentes conformations natives du système accélère la transition de ces conformations à une forme dénaturée de la protéine, ce qui, dans un contexte d'essai de liaison (cellules mortes, stress mécanique, température pièce), peut faire diminuer la concentration de protéine capable de lier l'apeline, donc également la radioactivité des filtres. Le fait que la radioactivité maximale dépend de la concentration de récepteurs signifie que la faible expression du récepteur sur la surface de la cellule diminue l'intensité de la radioactivité.

Le changement de constante de dissociation s'appuie sur la dynamique des conformations du système hAPJ-apeline. Deux phénomènes concomitants ont lieu dans un tel système : l'*induced fit* et la sélection conformationnelle. L'*induced fit* consiste en une modification de l'équilibre entre les conformations du récepteur par la liaison du ligand. La forme inactive de la protéine a une faible affinité pour le ligand et, une fois lié, celui-ci stabilise la forme active du récepteur, ce qui pousse l'équilibre vers la forme active liée du récepteur. La sélection conformationnelle consiste plutôt en la liaison privilégiée des formes actives du récepteur au ligand. La forme active d'un récepteur est également capable de lier le ligand avec une plus forte affinité que la forme inactive, donc la présence de ligand pousse l'équilibre des conformations vers la forme active liée du récepteur. Une déstabilisation de la forme active, avec ou sans ligand, est en mesure d'augmenter significativement la constante de dissociation du récepteur, puisque cette déstabilisation favorise la conformation inactive, dont l'affinité pour le ligand est faible.

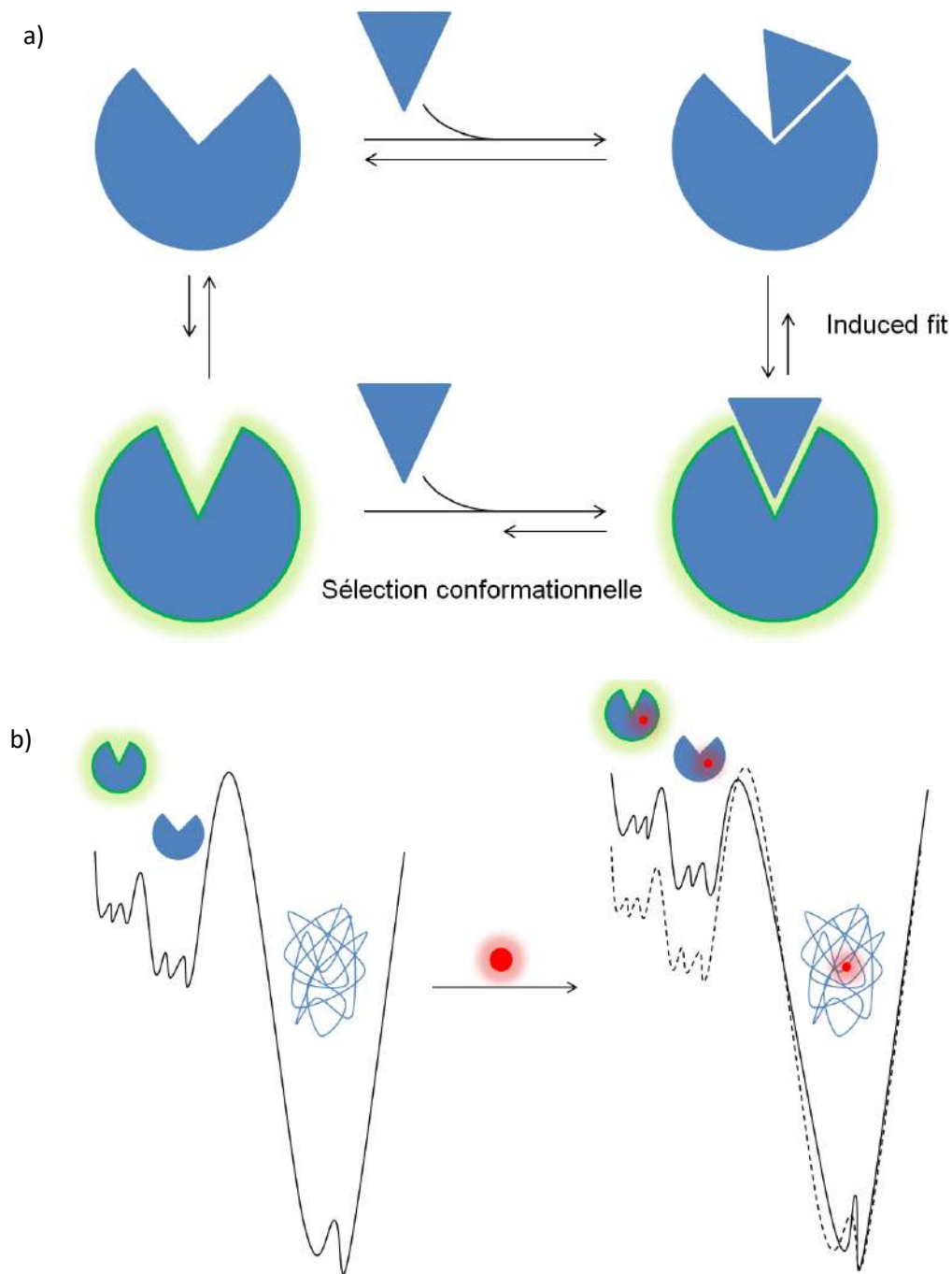


Figure 41 : Illustration des concepts d'*induced fit*, de sélection conformationnelle et d'instabilité induite par mutation

a) Illustration des concepts d'*induced fit* et de sélection conformationnelle. Le concept d'*induced fit* est la capacité d'un ligand à faciliter le changement conformationnel de sa protéine vers ses formes actives (ou inactives, selon le ligand). Le concept de sélection conformationnelle est plutôt la capacité d'un ligand à se lier préférentiellement aux formes actives de sa protéine, qui existent en équilibre avec les autres formes. Les deux concepts coexistent et constituent deux facettes du même phénomène qu'est l'activation d'une protéine par un ligand. b) Illustration du concept d'instabilité induite par mutation. Toute protéine existe en équilibre avec une ou plusieurs formes dénaturées, souvent plus stables que ses formes repliées. Cependant, la dénaturation peut, selon la protéine, être un processus lent. L'ajout d'une mutation (cercle rouge) peut déstabiliser davantage les formes repliées d'une protéine, accélérant ainsi le processus de dénaturation.

En comparant les données d'ELISA de surface des mutants quasi- ou non-liants à l'expression du WT de la même façon que pour le Mock, on constate que, à l'exception de F257G et Y309F, tous ces mutants ont un niveau d'expression différent et plus petit que celui du WT. Ceci expliquerait le faible signal observé et, par conséquent, la mauvaise qualité des résultats. Cependant, il existe aussi des mutants dont le niveau d'expression est faible, mais dont les données sont de meilleure qualité. Le faible niveau d'expression n'est donc pas une explication en elle-même. De même, tous les mutants ont été testés sous les mêmes conditions expérimentales, donc les erreurs expérimentales n'expliquent pas non plus la mauvaise qualité des données pour les mutants quasi- ou non-liants.

Chez les mutants non-liants, on sait que D75N et F257W ont le même niveau d'expression à la surface de la cellule que le WT. Le mutant W261R a un niveau d'expression plus élevé que le WT, tandis que les mutants N112W et S302N ont un niveau d'expression plus bas que le WT. On sait également que les mutants D75N, F257W et W261R ont un EC_{50} plus de dix fois plus élevé que le WT en BRET, autant pour $G\alpha_{i1}$ que pour la β -arrestine-2 et que le mutant N112W a un EC_{50} plus de dix fois plus élevé que le WT en essais BRET β -arrestine-2. La relation avec la constante de dissociation a été développée dans l'introduction (éqs 2,3) :

$$\tau = \frac{[R]_{tot}}{K_e}; \frac{K_D}{EC_{50}} = \frac{(1 + \tau)}{(1 + \alpha\tau)}; EC_{50} = \frac{K_D}{1 + \tau}, \alpha = 0$$

Bien que la mutation puisse avoir un impact sur le coefficient de transduction, en particulier ici sur K_e , ainsi que sur α , la combinaison d'une faible radioactivité maximale, d'une expression en surface égale ou supérieure au WT et d'une EC_{50} beaucoup plus élevée que le WT laisse supposer que le K_D a augmenté, au moins pour les mutants D75N, F257W et W261R (éq. 12). Pour le mutant N112W, cette explication reste plausible, mais insuffisante. Pour ce mutant en particulier, il est possible qu'une combinaison d'erreurs expérimentales, de plus faible expression en surface et d'un K_D légèrement augmenté suffisent à expliquer la qualité des résultats.

Dans le cas du mutant S302N, la qualité et la proximité des résultats de BRET à ceux du WT indiquent que la cause d'erreur, se situe certes dans son plus faible niveau d'expression, mais aussi possiblement dans le fait que cette mutation déstabilise le récepteur sous les conditions de liaison par compétition.

Chez les mutants quasi-liants, l'augmentation de la constante de dissociation est une explication peu plausible de la qualité des résultats. En effet, ces mutants n'ont aucune différence importante entre leurs EC_{50} et ceux du WT. Cependant, une combinaison des autres facteurs peut expliquer la qualité des données.

Analyse de la qualité et la validité des résultats des essais BRET

La qualité des essais BRET $G\alpha_{i1}$ effectués est faible, quoique suffisante. En effet, bien que la régression soit d'une exactitude suffisante pour la plupart des récepteurs, il n'en reste pas moins que l'incertitude sur les résultats reste significative et que les résultats d'une même expérience présentent une variation importante. Il en va de même pour les essais en β -arrestine-2, mais l'impact de cette variation des triplicatas d'une même expérience sont réduit. Ceci est dû au fait que la différence de rapport BRET pour l'essai $G\alpha_{i1}$ est fortement limitée, de l'ordre de 0,1, comparé au 0,9 des essais BRET β -arrestine-2, alors que l'écart type des triplicatas est sensiblement identique.

Cette différence d'amplitude de variation du rapport BRET est possiblement due à la localisation des protéines G et des arrestines avant et après l'activation du récepteur étudié. En effet, les protéines G sont généralement localisées à la membrane cytoplasmique sous leur forme inactive, puis les membres de la famille $G\alpha_{q/11}$ et $G\alpha_s$, mais pas $G\alpha_i$, passent de la membrane cytoplasmique au cytoplasme lors de leur activation (Wedegaertner, 2012). De même, les arrestines sont situées dans le cytoplasme, puis migrent vers la membrane cytoplasmique lorsque le récepteur devient phosphorylé, peu après son activation. Le lien entre localisation et magnitude de la différence de rapport BRET s'explique de deux façons : d'abord, la concentration effective des donneurs et accepteurs BRET, lorsque limitée à deux dimensions, est beaucoup plus élevée que lorsqu'elle peut diffuser dans trois dimensions; ensuite, le simple fait de se trouver contraint à deux dimensions, à concentration effective égale, augmente la probabilité de collision entre deux particules. Le

premier facteur s'explique par le fait que le volume effectif de la région autour de la membrane est bien inférieur au volume de la cellule entière. Le deuxième facteur, lui, est une conséquence directe du théorème de récurrence de Pólya (Doyle, 1982). Comme la probabilité de collision de deux protéines suivant un mouvement brownien augmente radicalement lorsque les protéines sont confinées à la membrane, la probabilité que les donneurs et accepteurs BRET se trouvent à l'intérieur du rayon critique nécessaire à l'observation de l'effet augmente, ce qui signifie que, dans le cas de $G\alpha_{i1}$, même si la sous-unité α est séparée du complexe $G\beta\gamma$, la probabilité de collision ou de proximité reste tellement élevée que la diminution du rapport BRET est faible.

Pour ce qui est de la validité des résultats des essais BRET, cela ne peut être évalué sans les résultats des ELISAs de surface. En effet, comme le niveau d'expression d'un mutant à la surface de la cellule a un impact sur son activité basale et l'activité maximale, ne pas considérer les données des ELISAs de surface peut mener à des erreurs. Par exemple en présumant qu'un mutant est deux fois plus exprimé que le type sauvage, mais avec des caractéristiques similaires, il pourrait être observé que le mutant possède une activité maximale deux fois plus élevée que le WT, alors que cette différence est due à une différence de niveau d'expression en surface (Schönege *et al*, 2017). Dans le cas des mutants qui nous intéressent le plus, soit D75N, N112G, D126V, V254L et W261R, les données d'ELISAs de surface indiquent que les mutants D75N et W261R, qui ont une activité maximale inférieure au WT tant en $G\alpha_{i1}$ qu'en arrestine, ont un niveau d'expression plus élevé que le WT, tandis que les mutants N112G, D126V et V254L, qui ont au moins une activité basale en $G\alpha_{i1}$ supérieure au WT, ont un niveau d'expression inférieur ou égal à celui du WT. Comme l'activité des mutants et leurs niveaux d'expression en surface ont des tendances inverses, deux conclusions sont possibles : l'activité constitutive des récepteurs mutants favorise leur internalisation, ce qui diminue leur expression apparente en surface, ou encore l'activité constitutive des récepteurs mutants est plus élevée que ce qui est observé. Dans les deux cas, le résultat d'activité constitutive reste véridique. Ce n'est pas la première fois que des mutants au profil fonctionnel biaisé sont identifiés (Ban *et al*, 2018), mais c'est la première fois que des mutants de hAPJ présentent une activité basale exclusive à une protéine effectrice.

Dans un autre ordre d'idées, il aurait été utile d'effectuer les essais avec comme référence à 0% le complexe $G\alpha_{i1}$ ou la β -arrestine-2 avec la CAAX-rGFP seuls, sans le récepteur. Sans cette information, il nous est impossible de déterminer le paramètre d'activité constitutive du récepteur de type sauvage. Pour rappel, nous avons pour le modèle opérationnel avec activité constitutive (éq. 3) :

$$\begin{aligned}\frac{E_0}{E_{max}} &= \frac{\alpha\tau}{(1 + \alpha\tau)} \\ \frac{E_\infty}{E_{max}} &= \frac{\tau}{(1 + \tau)} \\ \frac{K_D}{EC_{50}} &= \frac{(1 + \tau)}{(1 + \alpha\tau)}, \alpha \neq 1\end{aligned}$$

Ceci constitue un système à trois équations et trois inconnues : E_0 , E_∞ , K_D et EC_{50} peuvent normalement être obtenus via les essais de liaison par saturation, qui n'ont pas été effectués dans le cadre de ce travail, mais devront éventuellement être faits afin d'obtenir les paramètres α et τ , et la simple régression des essais BRET. E_{max} représente l'activité qu'aurait le récepteur si 100% de sa population était apte à signaler simultanément, $\alpha\tau$ représente la capacité relative du récepteur sans ligand de signaler et τ représente la capacité du récepteur avec ligand de signaler, le tout via la voie étudiée. Or, ici, sans la référence à 0% sans le récepteur, il est impossible de savoir où E_0 se trouve vraiment, car la seule véritable référence d'activité basale est celle du récepteur lui-même. Il s'ensuit que, sans une référence sans récepteur, le α du type sauvage est égal à zéro, car $E_0 = 0$ dans ce cas. Alors, nécessairement, toute mesure de changement relatif du paramètre α devient impossible à calculer. Par exemple, considérons le facteur σ du modèle opérationnel (éq. 5) :

$$\sigma = \log\left(\frac{\tau_{test}}{\tau_{ref}}\right)$$

L'équivalent pour α n'existe pas si $\alpha_{ref} = 0$. Ceci dit, un désavantage majeur du paramètre α est qu'il peut se trouver très près de 0, ce qui rend d'autant plus difficile son estimation, car dépendamment de l'écart type de la moyenne de trois observations, il est possible que la moyenne expérimentale du rapport BRET de la protéine G ou de l'arrestine seuls et ce même rapport avec le GPCR d'intérêt soient statistiquement identiques. Dans ce cas, à des fins de comparaison, il sera nécessaire d'assumer un α faible, mais présent. Par parcimonie, dans le cas des tests sur les mutants, la mesure assumée de α_{WT} devrait être celle calculée avec E_0 égal au maximum de l'intervalle de confiance de $E_{0;WT} - E_{0;blanc}$. De cette façon, l'activité constitutive d'un mutant serait identifiée dans un contexte où l'activité constitutive du récepteur WT est la plus haute envisageable avec le niveau de risque voulu.

Finalement, bien qu'une protéine G et une arrestine aient été utilisées pour étudier le profil de signalisation des différents mutants de hAPJ, il est impossible de généraliser ces conclusions à l'ensemble des protéines G et des arrestines, ni même aux autres protéines de la famille G_i ou la β -arrestine-1. Ces aspects de la signalisation des mutants générés ici n'ont pas été évalués; ils pourront l'être par ceux qui continueront ce projet.

Relation structure-activité des mutants hAPJ en relation à l'effet du sodium

Un aspect des GPCRs qui n'a pas été mentionné dans l'introduction est sa relation avec le sodium. Cette relation manque encore de clarté, puisque l'effet du sodium sur les GPCRs est encore mal caractérisé. Ce qui est connu est que les concentrations élevées de sodium ont un effet négatif sur la liaison d'agonistes aux GPCRs, sans effet pour les antagonistes (Pihlavisto *et al*, 1998; Schiffmann et Gimpl, 2018), qu'il existe un site de liaison « canonique » du sodium chez de nombreux GPCRs, centré autour du résidu D ou E2.50 (Vsevolod *et al*, 2014; Liu *et al*, 2012; White *et al*, 2018), et que certaines mutations imitant la présence du sodium dans son site de liaison ont un effet négatif autant sur la liaison d'agonistes que de l'activité des GPCRs (Schiffmann et Gimpl, 2018; White *et al*, 2018). De plus, un parallèle entre les GPCRs et les rhodopsines microbiennes, protéines à sept domaines transmembranaires servant entre-autres de pompes à ions activées par la

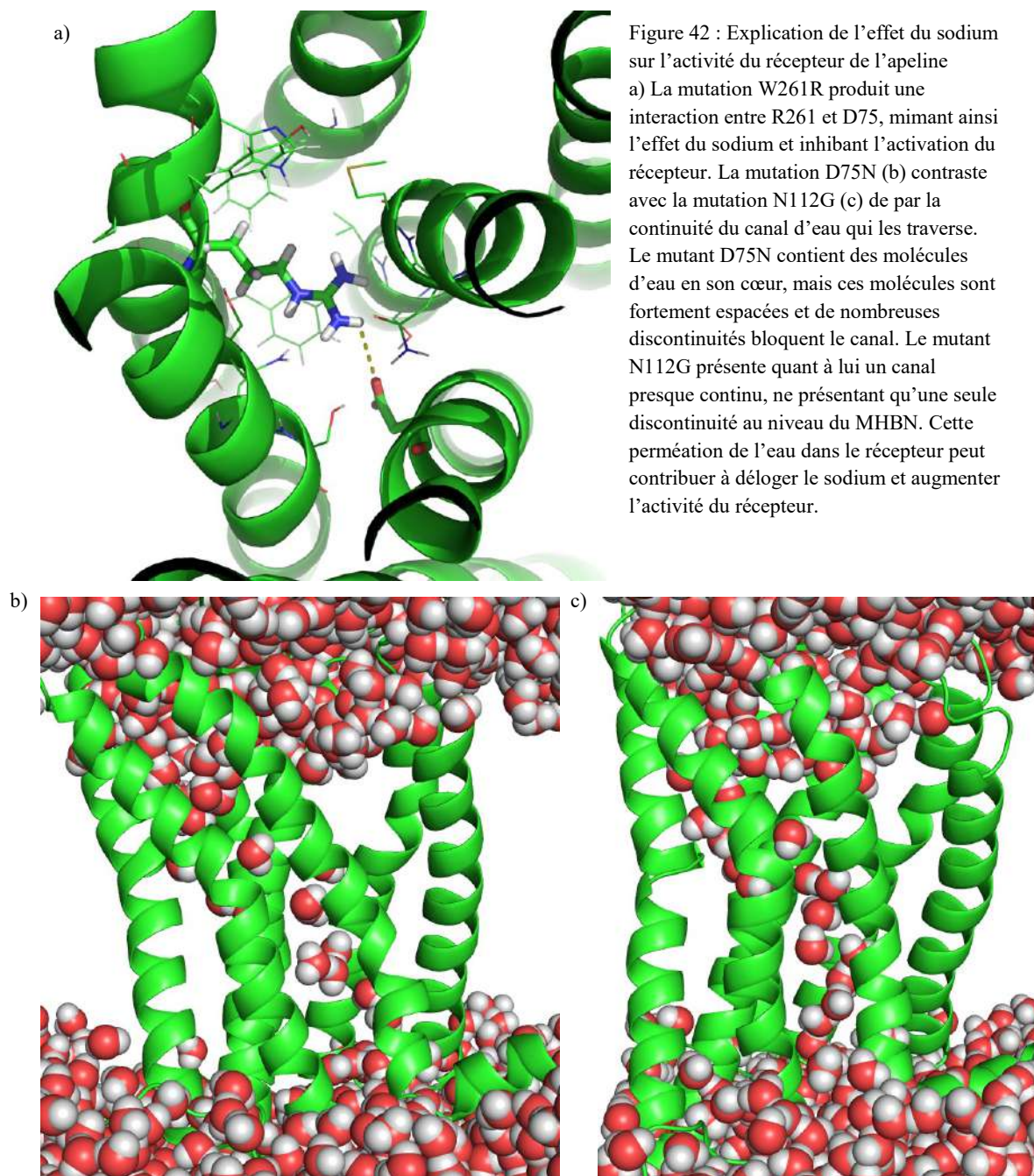
lumière, a été établi en fonction de la similitude de leurs sites de liaison du sodium (Shalaeva *et al*, 2015) et l'activité des GPCRs est influencée par la dépolarisation (Rinne *et al*, 2015). Ceci appuie l'hypothèse que les GPCRs ont évolué à partir de canaux ioniques et en conservent même une partie de la fonctionnalité (Vsevolod *et al*, 2014; Shalaeva *et al*, 2015).

L'analyse de la relation structure-activité des différents mutants de hAPJ du point de vue de l'effet du sodium peut principalement être concentrée sur les mutants D75N, N112G et W261R. Bien que les essais de liaison effectués sur les mutants D75N et W261R n'ont donné aucun résultat, cette absence apparente de signal peut s'expliquer par une perte d'affinité de l'apeline-13 radiomarquée pour le récepteur. Ceci dit, en l'absence de données sur la constante de dissociation, l'interprétation des résultats demeurera plus qualitative que quantitative pour ces deux mutants. Pour ce qui est du mutant N112G, son activité constitutive est confirmée et la liaison de l'apeline-13 au récepteur mutant l'est également.

Les trois mutants d'intérêt ont comme propriété commune qu'ils se trouvent au même endroit que le site de liaison attendu du sodium allostérique de hAPJ (Vsevolod *et al*, 2014; Liu *et al*, 2012; White *et al*, 2018). Considérant que le sodium a la capacité de réduire l'affinité des agonistes de nombreux GPCRs, sans impact sur les antagonistes, l'effet observé des mutants D75N et W261R sur les paramètres des essais de liaison et des essais BRET semble correspondre à celui du sodium. De plus, les deux mutations miment la présence du sodium à l'intérieur de hAPJ. En effet, la mutation D75N élimine la charge du résidu 2.50, ce qui imite la neutralisation de la charge de l'aspartate par le sodium. De même, la mutation W261R introduit un cation où se trouverait normalement le sodium, ce qui imite la présence du sodium en son site de liaison. Une illustration du concept se trouve à la figure 41.

L'effet du mutant N112G est l'inverse des mutants D75N et W261R. En effet, la mutation N112G élimine une partie du site de coordination du sodium. Cela peut ainsi diminuer l'affinité de ce site pour le sodium. Considérant le fait que la liaison du sodium ou un équivalent peut réduire non seulement l'activité maximale, mais également l'activité basale d'un récepteur si elle est présente (Schiffmann et Gimpl, 2018; White *et al*, 2018), on peut

inférer que l'effet du sodium ne se limite pas simplement à limiter l'affinité d'un récepteur pour certains ligands, mais contribue à diminuer la population des formes actives du récepteur, ce qui explique la diminution de l'activité basale et la sélectivité de l'effet du sodium sur les agonistes. Il est donc possible qu'une diminution de l'affinité de ce site pour le sodium augmente l'activité basale de hAPJ. Ceci expliquerait au moins en partie l'activité constitutive de hAPJ N112G. Une autre explication plausible de cette activité constitutive est que le mutant favorise l'ouverture d'un canal au milieu du cœur hydrophobe du GPCR, favorisant ainsi le passage du sodium du milieu extracellulaire au milieu intracellulaire (Vsevolod *et al*, 2014) et exploitant par le fait même le courant ionique issu du potentiel membranaire pour favoriser une conformation active du récepteur. En effet, de façon analogue à l'ATP synthase, le gradient de concentration du sodium peut être utilisé pour générer de l'énergie. Cette énergie peut être utilisée pour induire un changement conformationnel, dans ce cas vers une forme active du GPCR. Pour appuyer cette hypothèse, les positions des molécules d'eau des états des systèmes N112G apo et D75N apo ont été comparées. Dans la figure 41, il est montré que les formes les plus ouvertes du mutant D75N forment une colonne d'eau discontinue à plusieurs endroits, tandis que dans le cas du mutant N112G, certains états contiennent une colonne d'eau continue. En particulier, l'eau pénètre plus haut de façon continue dans la portion intracellulaire du mutant N112G. Cependant, il ne s'agit que d'une hypothèse. D'autres expériences devront être effectuées pour identifier les causes de l'activation constitutive de hAPJ par la mutation N112G.



Analyse comparative des systèmes mutants

Limitations de l'analyse théorique des différents systèmes de hAPJ

Tout comme les résultats expérimentaux, les résultats de dynamique moléculaire sont sujets à certaines limitations. Certaines de ces limitations ont été abordées dans l'introduction, mais d'autres restent à couvrir.

La première est le temps d'échantillonnage trop court. Considérons un événement rare dont chaque occurrence est indépendante des autres et se produisant λ fois par images, ou encore par seconde, alors le compte des occurrences de cet événement suit une loi de Poisson et la probabilité que cet événement se produise au moins une fois en k images (ou secondes) vaut $1 - e^{-k\lambda}$. Par conséquent, si on cherche le temps moyen nécessaire pour observer cet événement dans 95% des simulations, on obtient :

$$t = \frac{1}{\lambda} = \frac{k}{\ln(20)} \approx \frac{k}{3}$$

Considérant 10 μ s de simulation, ce temps vaut 3,34 μ s. Cela signifie que des changements conformationnels dont la fréquence moyenne est supérieure à un par 3,34 μ s ne sont pas observés au moins une fois avec un niveau de confiance de 95% dans le cadre de ce travail. Cela signifie également, dans le même ordre d'idées, que pour les événements ne se produisant aucune fois avec un niveau de confiance de 95% :

$$\frac{1}{\lambda} = \frac{k}{\ln(20) - \ln(19)} \approx 19.5k$$

On obtient ici à peu près 195 μ s. Le reste, entre les deux, est essentiellement sujet aux caprices du hasard. Il est important de mentionner que, considérant le fait que le calcul du taux d'occurrence moyen d'un événement doit être effectué avec au moins deux occurrences d'un événement, le temps de calcul requis pour mesurer le taux d'occurrence d'un événement de façon empirique (dans un contexte de dynamique moléculaire) est encore plus élevé.

La solution évidente est d'effectuer de plus longues simulations. Cependant, effectuer de plus longues simulations n'est pas une question de volonté ou de temps, mais également de ressources. Considérant le temps disponible, les ressources disponibles étaient insuffisantes pour effectuer davantage de calculs. Il s'ensuit que la longueur des simulations est plus courte et que la qualité des résultats en souffre. À l'heure actuelle, les ordinateurs les plus performants pour effectuer de la dynamique moléculaire sont les ordinateurs de la famille d'Anton (Shaw *et al*, 2008), qui disposent de *hardware* spécialisé qui accélère considérablement les calculs. Cependant, l'accès à ces ordinateurs spécialisés est limité par leur petit nombre. De décider si l'utilisation d'Anton est viable ou non est laissé à ceux qui poursuivront ce travail. Une autre option plus accessible est d'utiliser des superordinateurs dotés de processeurs graphiques. Le calcul des interactions non-covalentes est laissé au processeur graphique et le temps de calcul de chaque image s'en retrouve grandement réduit en conséquence. De plus longues simulations, ou davantage de simulations, pourront par conséquent être effectuées sur ce type de superordinateurs. Finalement, une approche utilisant le « coarse-graining » de l'eau et de la membrane, comme celle proposé par le champ de force ELBA (Orsi et Essex, 2011), aurait pu être utilisée. Pour ce faire, les molécules d'eau et les lipides sont simplifiés en leurs composantes stériques et leurs dipôles. Ainsi, l'eau devient une sphère et un dipôle, réduisant ainsi le nombre d'atomes requis pour représenter l'eau au tiers de la normale, et les lipides deviennent un ensemble de sphères et de dipôles ressemblant au lipide de départ, mais requérant seulement un peu plus du dixième d'atomes qu'il faudrait normalement utiliser pour les représenter. Cela diminue considérablement le nombre d'atomes de la simulation et améliore ainsi de beaucoup les performances, puisque le calcul des interactions entre atomes est dans $O(n^2)$.

La seconde est l'ensemble des approximations utilisées lors de la modélisation du système. D'abord, la membrane utilisée est un mélange de DOPC et de cholestérol, mais la réalité est beaucoup plus complexe et diversifiée dans une cellule normale. Non seulement une grande variété de lipides existent dans la membrane cytoplasmique, mais cette diversité est variable selon l'organe étudié et l'endroit observé de la cellule (Harayama et Riezman, 2018). Même en considérant la condition périodique aux limites en dynamique moléculaire, l'étude de l'effet des lipides sur un récepteur avec le niveau de précision et la

longueur de simulation utilisés dans le cadre de ce travail est techniquement impossible avec les moyens disponibles, puisqu'il serait nécessaire de simuler des systèmes considérablement plus grands que ceux étudiés ici pour inclure toute la diversité lipidique dans les différentes régions de la cellule. Comme les GPCRs baignent dans la membrane cytoplasmique, au beau milieu de tous ces lipides, il est raisonnable de penser que la composition de la membrane a un impact sur leur fonction. Il s'ensuit que les résultats observés ici sont des approximations, considérant une membrane idéalisée constituée de DOPC et de cholestérol. L'autre approximation est celle de l'eau. En effet, le modèle d'eau utilisé ici est le modèle SPC, qui ne contient que les trois atomes de la molécule d'eau avec les charges appropriées. Ce modèle est une grossière approximation de la véritable molécule d'eau, mais est utilisé ici parce que le champ de force GROMOS est optimisé avec ce modèle d'eau. D'autres modèles existent, mais la plupart sont moins simples, car ils incluent au moins un atome fictif supplémentaire pour mieux modéliser le comportement de l'eau. Parmi ceux-là, le modèle OPC est particulièrement intéressant puisque son optimisation n'est pas effectuée en fonction des contraintes géométriques de l'eau, hormis sa symétrie, mais plutôt en fonction de la proximité des propriétés physico-chimiques du modèle des valeurs expérimentales correspondantes (Izadi *et al*, 2014). Il pourrait être intéressant d'essayer quelques simulations avec ce modèle d'eau et vérifier les différences de dynamique qui pourraient apparaître.

La troisième limitation est la validité de la conformation de départ utilisée pour générer les ensembles hAPJ-WT-apo et hAPJ-WT-apeline. Cette limitation est partiellement compensée par la qualité du protocole d'échantillonnage, mais ne couvre pas tout. Si la conformation utilisée est trop éloignée d'une conformation réelle de hAPJ, l'échantillonnage effectué ici n'a aucune valeur. Ceci dit, il est peu probable que la structure obtenue par cristallographie soit si différente d'une structure de hAPJ-WT dans son milieu normal. Le danger se trouve plutôt dans l'interaction avec les protéines G. Nommément, il a précédemment été question de pré-complexes avec les protéines G (Ferré, 2015; Navarro *et al*, 2018). Si cette hypothèse vient à être confirmée, cela signifierait que la qualité l'échantillonnage effectué ici serait disparue ou sévèrement compromise, car cela changerait la dynamique du récepteur de façon importante. En effet,

les différentes interactions entre les protéines G et le récepteur limiteraient la mobilité de certaines régions du récepteur, altérant potentiellement la dynamique de ces régions. Cependant, comme l'hypothèse des pré-complexes reste une simple hypothèse, il n'était pas approprié de démarrer des simulations avec des pré-complexes et l'interprétation actuelle des résultats est plus près du consensus scientifique sur la dynamique des GPCRs.

Finalement, une dernière limitation des résultats de dynamique moléculaire tels que discutés ici est qu'il est toujours possible que la signalisation des GPCRs soit provoquée par l'exécution d'un changement conformationnel particulier. En effet, la formation du complexe GPCR-protéine G passe d'abord par la formation d'un complexe transitoire entre la protéine G et une forme « active-like » du récepteur (Gregorio *et al*, 2017). La nature de ce pré-complexe reste à ce jour mystérieuse et la formation du complexe GPCR-protéine G définitif pourrait impliquer non seulement des conformations particulières, tel qu'observé ici, mais également des changements conformationnels particuliers. Si ces changements conformationnels impliquent des conformations du GPCR exclusives aux systèmes constitutivement actifs ou avec ligands, le protocole d'analyse actuel serait capable d'identifier les points de départ et d'arrivée de ces changements, mais s'il s'agit de l'altération ou d'un nouveau changement conformationnel reliant deux conformations aussi observées dans des systèmes inactifs, alors la tICA ne pourra pas les repérer. En effet, la tICA est conçue pour repérer les changements conformationnels de faible fréquence, et elle le fait en repérant les états pour lesquels le temps de résidence des trajectoires (dans ces états) est le plus élevé. Elle ne révèle donc pas la nature des changements conformationnels, mais seulement leur finalité. Si l'on cherche à observer spécifiquement le déroulement du changement conformationnel observé, il faut plutôt utiliser les méthodes issues de la théorie des chemins de transition (*Transition-Path Theory*) (Weinan et Vanden-Eijnden, 2010). Ainsi, dans le cadre de ce travail, même si des conformations ont été identifiées comme conformes à la littérature pour des conformations capables d'activer des protéines G, la nature de cette activation reste en suspens et, si elle est de nature dynamique, les informations apportées ici ne pourront contribuer à la résolution de ce mystère sans être analysées de la bonne façon.

Une solution potentielle à ce problème, moyennant l'utilisation des méthodes de la *Transition-Path Theory*, est d'identifier les états pertinents du récepteur comme effectué ici, mais ensuite d'identifier les chemins de réaction soit dans l'espace conformationnel entier ou encore dans une forme étendue de celle utilisée pour identifier les états. Cela permettra d'éviter des chemins réactionnels différents qui auraient la même projection sur l'espace réduit. Une autre avenue intéressante de solution est de reproduire l'expérience de tICA/GHMM, mais uniquement sur les bouts de trajectoire qui font partie d'un état en particulier. Cela permettrait d'explorer la dynamique interne de cet état. Naturellement, cette exploration devra être faite sur l'espace conformationnel entier, car les changements à haute autocorrélation au sein d'un même état ne correspondent pas nécessairement à ceux du système entier.

Comparaison des systèmes hAPJ-WT-apeline et hAPJ-WT-apo

La comparaison des deux systèmes confirme la différence des systèmes hAPJ-WT-apeline et hAPJ-WT-apo. Le système hAPJ-WT-apeline présente des changements conformationnels au niveau de l'interface des TM3, 5 et 6, avec comme symptôme le changement conformationnel de F257, ainsi qu'au niveau de l'interface des TM1, 2 et 7, avec comme symptôme le réarrangement de la structure hélicale du TM7. Plus spécifiquement, ces changements impliquent soit les deux interfaces, soit celle des TM1, 2 et 7.

On sait que les composés biaisés vers les protéines G semblent perturber l'environnement intracellulaire des TM6 et 7, tandis que les composés biaisés vers les β -arrestines semblent perturber uniquement l'environnement intracellulaire du TM7 (Wacker *et al*, 2013). Par conséquent, il est raisonnable de supposer que le changement conformationnel qui consiste en un réarrangement de F257 ainsi que d'une distorsion de l'hélice du TM7 donne lieu à des conformations qui activent préférentiellement les protéines G, tandis que celui qui consiste en une dénaturation locale du milieu du TM7 donne lieu à des conformations qui lient préférentiellement via les β -arrestines.

Par contraintes temporelles, il a été décidé que la majorité des mutants testés allaient cibler l'activation des protéines G, puisque la littérature est plus riche à ce sujet que pour les β -arrestines. Cependant, avec rétrospective, quelques mutants (ou combinaisons de mutants) supplémentaires auraient pu être testés pour cibler leur activation, avec les plus prometteurs en caractères gras :

- **N46Q** : déstabilise la structure hélicale du TM7 par encombrement stérique
- D75E : idem à N46Q
- **S302P** ou S302G : déstabilise la structure hélicale au site observé de la déformation du TM7
- C303P ou C303G : idem à S302P ou S302G.
- Mêmes mutations jusqu'à N305.

L'objectif de ces mutations est de modifier la région située entre N46, D75, S302 et N305, de façon à perturber le TM7 sans perturber le TM6 et ainsi espérer pouvoir obtenir un mutant dont la signalisation s'effectue préférentiellement via les arrestines.

Analyse comparative des systèmes hAPJ-WT-apo, hAPJ-WT-apeline et hAPJ mutants

L'analyse de l'ensemble des distributions d'angles dièdres et des conformations adoptées par différents systèmes est une tâche complexe qu'il est approprié de subdiviser en plusieurs comparaisons avant de faire la synthèse de l'ensemble. Il est aussi important de souligner le fait que la majorité des conformations de l'ensemble des systèmes sont similaires entre-elles, avec quelques états minoritaires qui divergent significativement du reste. Les différences entre les systèmes seront donc basées sur les différences entre ces états minoritaires.

La première comparaison à effectuer est celle des systèmes hAPJ-D75N-apo et hAPJ-WT-apo, de façon à identifier les caractéristiques des formes inactives de hAPJ. Dans la figure 32, on peut voir deux changements conformationnels importants : un éloignement du TM5 des TM3 et 6, ainsi qu'un renforcement de l'interface entre les TM2, 3 et 4. L'effet du renforcement de l'interface des TM2, 3 et 4 est difficile à identifier. En effet, le TM4 est surtout connu pour servir d'interface à la dimérisation de certains GPCRs (Gorinski *et al*, 2012; Lee *et al*, 2003), et les mutations sur le TM4 ont un effet varié sur l'activité des

GPCRs (Vaisse *et al*, 2000; Décaillot *et al*, 2003; Liu *et al*, 2007; Wang, Tao, 2011; Rhee *et al*, 2008; Zhou *et al*, 2019). Cependant, considérant qu'il s'agit du mutant D75N et que cette conformation est unique à ce mutant, on peut supposer que cette interface sert de contrepoint aux interfaces des TM1,2 et 7, ainsi que des TM3, 5 et 6, en inhibant leurs changements conformationnels respectifs. Le découplage de la partie extracellulaire du TM5 vient compléter le tableau en désolidarisant les TM3, 5 et 6.

Vient ensuite la comparaison des systèmes hAPJ-WT-apeline et hAPJ-N112G-apo. Avec la figure 34, la correspondance entre les états des deux systèmes parle d'elle-même : les mêmes changements conformationnels du résidu F257 et du TM7 sont présents et combinés de la même façon dans les deux systèmes. Comme ces conformations ne sont présentes ni chez le mutant D75N, ni la forme Apo du WT, on peut en conclure qu'il s'agit de formes actives de hAPJ. Une observation particulière qui vient apporter un bémol à cette conclusion est la figure 35 a), où la distribution de l'angle dièdre pour le mutant D75N laisse présumer que le résidu F257 oscille entre sa conformation active et sa conformation inactive.

Considérant que les autres paramètres conformationnels étudiés ne correspondent pas à ceux du mutant N112G, ni de ceux du système WT-apeline, deux explications sont plausibles :

- La réorganisation des différentes interfaces du récepteur rend inopérant le changement conformationnel de F257
- Ce changement conformationnel est un symptôme, ou encore une cause facultative, de l'activation du récepteur plutôt qu'une cause nécessaire.

La comparaison du système hAPJ-WT-apeline avec les systèmes hAPJ-D126V-apo (Figure 36) et hAPJ-V254L-apo (Figure 38) apporte des informations qui font pencher la balance vers la deuxième possibilité. En effet, autant le mutant D126V que le mutant V254L ont une activité constitutive sur la voie $G\alpha_{i1}$, mais les conformations de ces deux systèmes ne présentent pas de changement conformationnel de F257. Cela signifie que ce changement conformationnel n'est pas obligatoire à la signalisation par $G\alpha_{i1}$, sans pour autant invalider la possibilité que le changement conformationnel de F257 soit une cause alternative de ce

type de signalisation. Ce qui est aussi surprenant est que les mutants D126V et V254L semblent tous deux avoir une façon distincte de provoquer la signalisation via les protéines G : la mutation D126V facilite la dissociation du TM3 du TM6 en augmentant sa propension à interagir avec les TM2 et 4 (Figure 36), tandis que la mutation V254L perturbe l'interaction entre les TM5 et 6 par encombrement stérique (Figure 38). Si ces états sont bel et bien ceux qui ont la capacité d'activer $G\alpha_{i1}$, ce qui est plausible considérant la littérature existante, cela signifie que l'activation de cette protéine ne dépend pas d'un changement conformationnel particulier, mais bien d'une déstabilisation de l'interface formée par les TM3, 5 et 6, peu importe le moyen utilisé. Considérant que la finalité du changement conformationnel de cette interface implique une dissociation significative des TM3 et 6 et que le résidu F257 n'interagit pas avec les protéines G, cette explication semble logique.

Pour ce qui est de l'interaction avec les arrestines, les résultats ne sont malheureusement pas clairs. En effet, comme les seuls systèmes analysés ayant une activité sur la voie de la β -arrestine-2 exhibent également une activité pour la voie $G\alpha_{i1}$, il est impossible de dire si la perturbation de l'interface des TM1, 2 et 7 est un signe précurseur de l'activation de la voie des arrestines ou si ce n'est qu'un autre symptôme de l'activation de la voie des protéines G via une conformation différente. Ceci dit, considérant le fait que la signalisation via les arrestines implique principalement le TM7 et que les mutants ayant une activité constitutive pour $G\alpha_{i1}$ n'impliquent aucunement le TM7, il peut être raisonnable d'inférer que cette perturbation correspond à une capacité à activer la voie des arrestines. Cela correspondrait aux résultats observés dans la littérature sur le sujet (voir « Structure et signalisation biaisée » dans l'introduction). Si ce constat est vrai, alors les mutations suggérées dans la section précédente ont le potentiel de provoquer une signalisation constitutive biaisée vers la β -arrestine-2.

Ainsi, les résultats de ce travail viennent appuyer les observations effectuées sur le récepteur β_2 -adrénergique (Liu *et al*, 2012; Wacker *et al*, 2013), ainsi que les différentes structures de GPCRs en complexe avec une protéine G hétérotrimérique. Cependant, les résultats obtenus ne peuvent éclairer sur la nature des changements conformationnels

essentiels à l'activation de la voie des arrestines. Ils sont non-concluants, même s'ils semblent concorder avec les résultats expérimentaux existants.

Conclusion

Bien que les résultats de ce travail ne parviennent pas à élucider le mystère de la relation entre structure et signalisation biaisée des GPCRs, ils apportent néanmoins des informations pertinentes à sa résolution. D'abord, le protocole d'échantillonnage et la technique d'analyse par sous-matrices principales développés ici ont permis une exploration d'une partie conséquente de l'espace conformationnel des différents systèmes dérivés d'hAPJ, ainsi que d'identifier différents changements conformationnels qui pourraient être responsables de sa signalisation normale. Ce protocole, bien qu'il requière encore des améliorations substantielles, a le potentiel de devenir une méthode hautement efficace d'exploration et d'analyse de l'espace conformationnel des protéines par la dynamique moléculaire. Les résultats de ce protocole ont révélé, en comparant les systèmes hAPJ-WT-apo et hAPJ-WT-apeline, que la région autour des résidus F257^{6,44}, W261^{6,48}, ainsi que du motif NPxxY, est impliquée dans l'activation du récepteur.

Ces résultats ont à leur tour orienté en partie le design de mutants, dont cinq, D75N, N112G, D126V, V254L et W261R, présentent un profil fonctionnel substantiellement différent de celui du hAPJ WT. Deux de ces mutants, D75N et W261R, ont une puissance et une efficacité d'activation par l'apeline diminuée par rapport au WT. Le mutant N112G présente une activité basale importante, autant sur la voie $G\alpha_{i1}$ que la voie β -arrestine 2, tandis que les mutants D126V et V254L présentent une activité basale sur la voie $G\alpha_{i1}$ uniquement. Ces trois mutants en particulier constituent une première occurrence connue de mutants de hAPJ présentant une importante activité basale, dans le cas de N112G, et dont l'activité basale est exclusive à une seule des deux principales voies de signalisation du récepteur, dans le cas de D126V et V254L.

La simulation et l'analyse du système hAPJ-N112G-apo a permis de confirmer que les régions identifiées en comparant les systèmes hAPJ-WT-apo et hAPJ-WT-apeline sont bien impliquées dans l'activation du récepteur, tandis que les mutants D126V et V254L ont élargi le constat dans le cas spécifique de l'activation de la voie $G\alpha_{i1}$. En effet, les systèmes hAPJ-D126V-apo et hAPJ-V254L-apo présentent tous deux une activité exclusive à la voie $G\alpha_{i1}$, mais les manifestations conformationnelles sont différentes, ce qui indique que la

déstabilisation de l'interface des TM3, 5 et 6 semble être le seul déterminant structural de l'activation de la voie $G\alpha_{i1}$ par le récepteur.

Tous ces résultats, aussi intéressants et informatifs soient-ils, donnent lieu à de nombreuses questions restées sans réponses. Beaucoup reste à faire : le protocole d'analyse de simulations et les outils connexes développés ici doivent encore être éprouvés sur des systèmes bien connus afin d'évaluer correctement leur qualité; les paramètres α et τ des mutants les plus intéressants devront éventuellement être calculés et les plus menus détails de leur profil fonctionnel devront être étudiés; la relation des mutants d'intérêt avec l'effet du sodium, en particulier D75N, W261R et N112G, devra être évaluée avec plus de détails; l'aspect structural de la signalisation biaisée de hAPJ devra être étudié à l'aide de techniques expérimentales... De même, ces résultats n'expliquent pas tout, en particulier en ce qui concerne la signalisation biaisée vers les β -arrestines. Un mutant au même profil que D126V ou V254L, mais dont l'activité est exclusive à la β -arrestine 2 se fait toujours attendre. Malgré ces faiblesses, des avenues de recherches restent accessibles et certaines d'entre-elles sont mentionnées dans la discussion. De plus, il n'en reste pas moins que l'ensemble de ces résultats, en plus de faire progresser substantiellement les connaissances sur le comportement du récepteur de l'apeline et des GPCRs en général, fournit de nombreux outils de recherche pour de futures études sur le récepteur, ainsi que, plus largement, pour l'étude de simulations de dynamique moléculaire d'autres systèmes.

Liste des références

Abraham MJ *et al* (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 12:19-25

Abraham MJ *et al* (2018) *GROMACS Reference Manual : Version 2018.2*. <http://manual.gromacs.org/>, page consultée le 29 août 2019

Andrews DL (2009) Resonance Energy Transfer : Theoretical Foundations and Developing Applications. Dans Noginov MA *et al*. *Tutorials in Complex Photonic Media*. SPIE (1e éd. 2009). p461-499

Bakker RA *et al* (2008) Constitutively Active Mutants of the Histamine H1 Receptor Suggest a Conserved Hydrophobic Asparagine-Cage That Constrains the Activation of Class A G Protein-Coupled Receptors. *Mol Pharmacol* 73(1):94-103

Ballesteros J, Weinstein H (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci* 25(C):366-428

Ballesteros JA *et al* (2001) Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J Biol Chem* 276(31):29171-29177

Ban T *et al* (2018) GPCR structure and function relationship: identification of a biased apelin receptor mutant. *Biochem J* 475(23):3813-3826

Baranski TJ *et al* (1999) C5a Receptor Activation Genetic Identification of Critical Residues in Four Transmembrane Helices. *J Biol Chem* 274(22):15757-15765

Beauchamp KA *et al* (2011) MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J Chem Theory Comput* 7(10):3412-3419

Berendsen HJC *et al* (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Simulations* 91(1-3):43-56

Berman DM, Gilman AG (1998) Mammalian RGS Proteins: Barbarians at the Gate. *J Biol Chem* 273(3):1269-1272

Besserer-Offroy E *et al* (2018) The hypotensive effect of activated apelin receptor is correlated with β -arrestin recruitment. *Pharmacol Res* 131:7-16

Bhattacharyya A (1943) On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distributions. *Bulletin of the Calcutta Mathematical Society* 35:99-109

- Bishop CM (2006) Pattern Recognition and Machine Learning. Springer Series in Information Science and Statistics (1e éd. 2006). p1-120 et p603-635
- Black JW, Leff P (1983) Operational models of pharmacological agonism. *Proc R Soc London* 220:141-162
- Bohn LM *et al* (2000) Mu-opioid receptor desensitization by beta-arrestin-2 determines morphine tolerance but not dependence. *Nature* 408:720-723
- Bonde MM *et al* (2010) Biased Signaling of the Angiotensin II Type 1 Receptor Can Be Mediated through Distinct Mechanisms. *PLoS One* 5(11):e14135
- Brame AL *et al* (2015) Design, characterization, and first-in-human study of the vascular actions of a novel biased apelin receptor agonist. *Hypertension* 65:834-840
- Bridges TM, Lindsley CW (2008) G-Protein-Coupled Receptors: From Classical Modes of Modulation to Allosteric Mechanisms. *ACS Chem Biol* 3(9):530-541
- Brizzard B (2008) Epitope tagging. *BioTechniques* 44(5):693-695
- Burger M *et al* (1999) Point Mutation Causing Constitutive Signaling of CXCR2 Leads to Transforming Activity Similar to Kaposi's Sarcoma Herpesvirus-G Protein-Coupled Receptor. *J Immunol* 163(4):2017-2022
- Cabana J *et al* (2012) Critical Hydrogen Bond Formation for Activation of the Angiotensin II Type 1 Receptor. *J Biol Chem* 288(4):2593-2604
- Cabana J *et al* (2015) Identification of Distinct Conformations of the Angiotensin-II Type 1 Receptor Associated with the Gq/11 Protein Pathway and the β -Arrestin Pathway Using Molecular Dynamics Simulations. *J Biol Chem* 290(25):15835-15854
- Carpenter B *et al* (2016) Structure of the adenosine A(2A) receptor bound to an engineered G protein. *Nature* 536:104-107
- Ceraudo E *et al* (2014) Biased signaling favoring gi over β -arrestin promoted by an apelin fragment lacking the C-terminal phenylalanine. *J Biol Chem* 289(35):24599-24610
- Chagnon F *et al* (2017) ELABELA Improves Cardio-Renal Outcome in Fatal Experimental Septic Shock. *Crit Care Med* 45(4):pe391-e398
- Chen A *et al* (2001) Constitutive Activation of A3 Adenosine Receptors by Site-Directed Mutagenesis. *Biochem Biophys Res Commun* 284(3):596-601
- Cheng Y, Prusoff WH (1973) Relationship between the inhibition constant (K₁) and the concentration of inhibitor which causes 50 per cent inhibition (I₅₀) of an enzymatic reaction. *Biochemical Pharmacology* 22(23):3099-3108

Cheung WY (1980) Calmodulin plays a pivotal role in cellular regulation. *Science* 207:19-27

Choi E, Lee C (2003) Feature extraction based on the Bhattacharyya distance. *Pattern Recognition* 36:1703-1709

Clegg RM (2008) Förster resonance energy transfer — FRET what is it, why do it, and how it's done. Dans T.W.J. Gadella (Éd). *FRET and FILM Techniques*. Elsevier (1e éd. 2008). p1-57

Coquerel D *et al* (2018) The apelinergic system as an alternative to catecholamines in low-output septic shock. *Crit Care* 22(10):1-7

Coveney PV, Wan S (2016) On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys Chem Chem Phys* 18:30236-30240

Doddi H *et al* (2019) Exact Topology Learning in a Network of Cyclostationary Processes. *American Control Conference*, <https://arxiv.org/pdf/1903.09210.pdf>, page consultée le 29 août 2019

Doyle PG (1982) *Application of Rayleigh short-cut method to Pólya's recurrence theorem*. Dartmouth College, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.12.1402&rep=rep1&type=pdf>, page consultée le 29 août 2019.

Draper-Joyce CJ *et al* (2018) Structure of the adenosine-bound human adenosine A1 receptor-Gi complex. *Nature* 558:559-563

Dror RO *et al* (2011) Activation mechanism of the β 2-adrenergic receptor. *PNAS* 108(46):18684-18689

Décaillot FM *et al* (2003) Opioid receptor random mutagenesis reveals a mechanism for G protein-coupled receptor activation. *Nat Struct Biol* 10(8):629-636

D'Antona AM *et al* (2006) A cannabinoid receptor 1 mutation proximal to the DRY motif results in constitutive activity and reveals intramolecular interactions involved in receptor activation. *Brain Res* 1108(1):1-11

D'Antona AM *et al* (2006) Mutations of CB1 T210 Produce Active and Inactive Receptor Forms: Correlations with Ligand Affinity, Receptor Stability, and Cellular Localization. *Biochemistry* 45(17):5606-5617

El Messari S *et al* (2004) Functional dissociation of apelin receptor signaling and endocytosis: implications for the effects of apelin on arterial blood pressure. *J Neurochem* 90:1290-1301

- Elgeti M *et al* (2013) Precision vs Flexibility in GPCR signaling. *JACS* 135:12305-12312
- Fan X *et al* (2003) Structural and Functional Study of the Apelin-13 Peptide, an Endogenous Ligand of the HIV-1 Coreceptor, APJ. *Biochemistry* 42:10163-10168
- Fanelli F (2000) Theoretical study on mutation-induced activation of the luteinizing hormone receptor. *J Mol Biol* 296(5):1333-1351
- Fang Y *et al* (2015) Orphan GPCRs As Emerging Drug Targets. *Front Pharmacol* 6(295):1-2
- Feng W, Song ZH (2003) Effects of D3.49A, R3.50A, and A6.34E mutations on ligand binding and activation of the cannabinoid-2 (CB2) receptor. *Biochem Pharmacol* 65(7):1077-1085
- Feng Y *et al* (1998) Mechanism of Constitutive Activation of the AT1 Receptor: Influence of the Size of the Agonist Switch Binding Residue Asn111. *Biochemistry* 37(45):15791-15798
- Ferris CD, Snyder SH (1992) Inositol 1,4,5-Trisphosphate-Activated Calcium Channels. *Annu Rev Physiol* 54:469-488
- Ferré S (2015) The GPCR heterotetramer: challenging classical pharmacology. *Trends Pharmacol Sci* 36(3):145-152
- Fredriksson R *et al* (2003) The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol Pharmacol* 63(6):1256-1272
- Fredriksson R, Schiöth HB (2005) The repertoire of G-protein coupled receptors in fully sequenced genomes. *Mol Pharmacol* 67(5):1414-1425
- Gainetdinov RR *et al* (2004) Desensitization of G Protein–Coupled Receptors and Neuronal Functions. *Annu Rev Neurosci* 27:107-144
- Galés C *et al* (2006) Probing the activation-promoted structural rearrangements in preassembled receptor–G protein complexes. *Nat Struct Mol Biol* 13:778-786
- García-Nafria J *et al* (2018) Cryo-EM structure of the adenosine A2A receptor coupled to an engineered heterotrimeric G protein. *eLife* 7:e35946
- Gardinier C (2009) *Stochastic Methods : A Handbook for the Natural and Social Sciences*. Springer Series in Synergetics (4e éd. 2009)

- Gerbier R *et al* (2015) New structural insights into the apelin receptor: identification of key residues for apelin binding. *FASEB J* 29:314-322
- Gorinski N *et al* (2012) Computational and Experimental Analysis of the Transmembrane Domain 4/5 Dimerization Interface of the Serotonin 5-HT_{1A} Receptor. *Mol Pharmacol* 82(3):448-463
- Greasley PJ *et al* (2002) Mutagenesis and Modelling of the α 1b-Adrenergic Receptor Highlight the Role of the Helix 3/Helix 6 Interface in Receptor Activation. *Mol Pharmacol* 61(5):1025-1032
- Gregorio GG *et al* (2017) Single-molecule analysis of ligand efficacy in β 2AR-G-protein activation. *Nature* 547:68-73
- Gundry J *et al* (2017) A Practical Guide to Approaching Biased Agonism at G Protein Coupled Receptors. *Front Neurosci* 11(17):1-6
- Han M *et al* (2001) Crystal Structure of β -Arrestin at 1.9 Å: Possible Mechanism of Receptor Binding and Membrane Translocation. *Structure* 9(9):869-880
- Harayama T, Riezman H (2018) Understanding the diversity of membrane lipid composition. *Nat Rev Mol Cell Biol* 19(5):281-296
- Harrigan MP *et al* (2017) MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys J* 112(1):10-15
- Hepler JR, Gilman AG (1992) G proteins. *Trends Biochem Sci* 17(10):393-387
- Ho KC, Hamelberg D (2016) Oscillatory Diffusion and Second-Order Cyclostationarity in Alanine Tripeptide from Molecular Dynamics Simulation. *J Chem Theory Comput* 12:372-382
- Ho SN *et al* (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77(1):51-59
- Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discovery* 1:727-730
- Huynh J *et al* (2009) Role of helix 8 in G protein-coupled receptors based on structure–function studies on the type 1 angiotensin receptor. *Mol and Cell Endocrinol* 302:118-127
- Hynes RO (2009) The Extracellular Matrix : Not Just Pretty Fibrils. *Science* 326:1216
- Hünenberger PH (2005) Thermostat Algorithms for Molecular Dynamics Simulations. *Adv Polym Sci* 173:105-149

- Ibsen MS *et al* (2019) Cannabinoid CB1 and CB2 Receptor-Mediated Arrestin Translocation: Species, Subtype, and Agonist-Dependence. *Front Pharmacol* 10(350):1-18
- Izadi S *et al* (2014) Building Water Models: A Different Approach. *J Phys Chem Lett* 5:3863-3871
- Kang Y *et al* (2013) Apelin-APJ signaling is a critical regulator of endothelial MEF2 activation in cardiovascular development. *Circ Res* 113:22-31
- Kang Y *et al* (2015) Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* 523:561-567
- Katrach V *et al* (2013) Structure-function of the G protein-coupled receptor superfamily. *Annu Rev Pharmacol Toxicol* 53(1):531-556
- Kenakin T (2011) Functional Selectivity and Biased Receptor Signaling. *J Pharmacol Exp Ther* 336:296-302
- Kenakin T, Christopoulos A (2013) Signalling bias in new drug discovery: detection, quantification and therapeutic impact. *Nat Drug Disc* 12:205-216
- Kim J (2014) Apelin-APJ Signaling: a Potential Therapeutic Target for Pulmonary Arterial Hypertension. *Mol Cells* 37(3):196-201
- Kim TH *et al* (2013) The Role of Ligands on the Equilibria Between Functional States of a G Protein-Coupled Receptor. *JACS* 135(25):9465-9474
- Klein Herenbrink C *et al* (2016) The role of kinetic context in apparent biased agonism at GPCRs. *Nat Comm* 7:art. 10842
- Koehl A *et al* (2018) Structure of the μ -opioid receptor-Gi protein complex. *Nature* 558:547-552
- Kofuku Y *et al* (2012) Efficacy of the β 2-adrenergic receptor is determined by conformational equilibrium in the transmembrane region. *Nat Comm* 3:art. 1045
- Kolodny A *et al* (2015) The Prescription Opioid and Heroin Crisis: A Public Health Approach to an Epidemic of Addiction. *Annu Rev Public Health* 36:559-574
- Komolov KE *et al* (2017) Structural and Functional Analysis of a β 2-Adrenergic Receptor Complex with GRK5. *Cell* 169:407-421
- Laskowski RA *et al* (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283-291

Latorraca NR *et al* (2018) Molecular mechanism of GPCR-mediated arrestin activation. *Nature* 557:452-456

Lee DK *et al* (2005) Modification of the terminal residue of apelin-13 antagonizes its hypotensive action. *Endocrinology* 146(1):231-236

Lee SP *et al* (2003) D2 Dopamine Receptor Homodimerization Is Mediated by Multiple Sites of Interaction, Including an Intermolecular Interaction Involving Transmembrane Domain 4. *Biochemistry* 42(37):11023-11031

Levin DA *et al* (2008) *Markov Chains and Mixing Times*. American Mathematical Society (1e éd. 2008)

Li J *et al* (2001) Constitutive Activation of the μ Opioid Receptor by Mutation of D3.49(164), but Not D3.32(147): D3.49(164) Is Critical for Stabilization of the Inactive Form of the Receptor and for Its Expression. *Biochemistry* 40(40):12039-12050

Liang YL *et al* (2017) Phase-plate cryo-EM structure of a class B GPCR–G-protein complex. *Nature* 546:118-123

Liang YL *et al* (2018) Phase-plate cryo-EM structure of a biased agonist-bound human GLP-1 receptor-Gs complex. *Nature* 555:121-125

Liggett SB (2011) Phosphorylation Barcoding as a Mechanism of Directing GPCR Signaling. *Sci Signaling* 4(185):pe36

Liu G *et al* (2007) Four Missense Mutations in the Ghrelin Receptor Result in Distinct Pharmacological Abnormalities. *J Pharmacol Exp Ther* 322(3):1036-1043

Liu JJ *et al* (2012) Biased signaling pathways in β 2-adrenergic receptor characterized by 19F-NMR. *Science* 335:1106-1110

Liu W *et al* (2012) Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions. *Science* 337(6091):232-236

Luttrell LM *et al* (2002) The role of β -arrestins in the termination and transduction of G-protein-coupled receptor signals. *J Cell Sci* 115:455-465

Luttrell LM, Gesty-Palmer D (2010) Beyond Desensitization: Physiological Relevance of Arrestin-Dependent Signaling. *Pharmacol Rev* 62(2):305-330

Ma Y *et al* (2017) Structural Basis for Apelin Control of the Human Apelin Receptor. *Structure* 25(6):858-866(e4)

Mahalingam M *et al* (2008) Two protonation switches control rhodopsin activation in membranes. *PNAS* 105(46):17795-17800

- Manglik A *et al* (2016) Structure-based discovery of opioid analgesics with reduced side effects. *Nature* 537(7619):185-190
- Margathe J-F *et al* (2014) Structure Activity Relationship Studies toward the Discovery of Selective Apelin Receptor Agonists. *J Med Chem* 57:2908-2919
- Marie J *et al* (1999) Constitutive Activation of the Human Bradykinin B2 Receptor Induced by Mutations in Transmembrane Helices III and VI. *Mol Pharmacol* 55(1):92-101
- Masri B *et al* (2006) The apelin receptor is coupled to Gi1 or Gi2 protein and is differentially desensitized by apelin fragments. *J Biol Chem* 281:18317-18326
- McCorvy JD *et al* (2018) Structure-inspired design of β -arrestin-biased ligands for aminergic GPCRs. *Nat Chem Biol* 14:126-134
- McGibbon RT *et al* (2014) Understanding Protein Dynamics with L1-Regularized Reversible Hidden Markov Models. *Proceedings of the 31st International Conference on Machine Learning (ICML)*
- McGibbon RT *et al* (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* 109(8):1528-1532
- Medhurst AD *et al* (2003) Pharmacological and immunohistochemical characterization of the APJ receptor and its endogenous ligand apelin. *J Neurochem* 84(5):1162-1172
- Mirnic K *et al* (2001) Disease-specific changes in regulator of G-protein signaling 4 (RGS4) expression in schizophrenia. *Mol Psychiatry* 6:293-301
- Moore CAC *et al* (2007) Regulation of Receptor Trafficking by GRKs and Arrestins. *Annu Rev Physiol* 69:451-482
- Murza A *et al* (2012) Elucidation of the Structure–Activity Relationships of Apelin: Influence of Unnatural Amino Acids on Binding, Signaling, and Plasma Stability. *Chem Med Chem* 7:318-325
- Murza A *et al* (2015) C-Terminal Modifications of Apelin-13 Significantly Change Ligand Binding, Receptor Signaling, and Hypotensive Action. *J Med Chem* 58:2431-2440
- Murza A *et al* (2016) Discovery and Structure–Activity Relationship of a Bioactive Fragment of ELABELA that Modulates Vascular and Cardiac Functions. *J Med Chem* 59:2962-2972
- Murza A *et al* (2019) Apelins, ELABELA, and their derivatives: Peptidic regulators of the cardiovascular system and beyond. *Pep Sci* 111(1):e24064

- Murza A, Sainsily X *et al* (2017) Structure–activity relationship of novel macrocyclic biased apelin receptor agonists. *Org Biomol Chem* 15:449-458
- Namkung Y *et al* (2016) Monitoring G protein-coupled receptor and β -arrestin trafficking in live cells using enhanced bystander BRET. *Nat Comm* 7:art. 12178
- Navarro G *et al* (2018) Evidence for functional pre-coupled complexes of receptor heteromers and adenylyl cyclase. *Nat Comm* 9:art. 1242
- Neves SR *et al* (2002) G Protein Pathways. *Science* 296(5573):1636-1639
- Nickelsen D (2014) Markov Processes linking Thermodynamics and Turbulence. Institut für Physik, Carl von Ossietzky Universität Oldenburg. p5-104
- Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *J Chem Phys* 78:1055-1076
- Noé F, Nüske F (2013) A Variational Approach to Modeling Slow Processes in Stochastic Dynamical Systems. *Multiscale Model Simul* 11(2):635-655
- Nygaard R *et al* (2009) Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol Sci* 30(5):249-259
- Nygaard R *et al* (2013) The dynamic process of $\beta(2)$ -adrenergic receptor activation. *Cell* 152:532-542
- O’Carroll A *et al* (2013) The apelin receptor APJ: journey from an orphan to a multifaceted regulator of homeostasis. *J Endocrinol* 219:pR13-R35
- Orsi M, Essex JW (2011) The ELBA Force Field for Coarse-Grain Modeling of Lipid Membranes. *PLOS One* 6(12):e28637
- Palcewski K *et al* (1994) Structure and functions of arrestins. *Protein Sci* 3:1355-1361
- Park SH *et al* (2012) Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* 491:779-784
- Parnot C *et al* (2000) Systematic identification of mutations that constitutively activate the angiotensin II type 1A receptor by screening a randomly mutated cDNA library with an original pharmacological bioassay. *PNAS* 97(13):7615-7620
- Pfleger KDG, Eidne KA (2006) Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat Methods* 3(3):165-174
- Pihlavisto M *et al* (1998) Modulation of agonist binding to recombinant human K2-adrenoceptors by sodium ions. *Biochim Biophys Acta* 1448:135-146

- Pitcher JA *et al* (1998) G Protein–Coupled Receptor Kinases. *Annu Rev Biochem* 67:653-692
- Premont RT, Gainetdinov RR (2007) Physiological Roles of G Protein–Coupled Receptor Kinases and Arrestins. *Annu Rev Physiol* 69:511-534
- Pérez-Hernández G *et al* (2013) Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 139(15102):1-13
- Quoyer J *et al* (2013) Pepducin targeting the C-X-C chemokine receptor type 4 acts as a biased agonist favoring activation of the inhibitory G protein. *PNAS* 110(52):E5088-E5097
- Raehal KM *et al* (2005) Morphine side effects in beta-arrestin 2 knockout mice. *J Pharmacol Exp Ther* 314(3):1195-1201
- Rajagopal S *et al* (2011) Quantifying Ligand Bias at Seven-Transmembrane Receptors. *Mol Pharmacol* 80(3):367-377
- Rasmussen SGF *et al* (2011) Crystal structure of the β 2 adrenergic receptor–Gs protein complex. *Nature* 477:549-557
- Rhee M *et al* (2008) Functional Role of Tryptophan Residues in the Fourth Transmembrane Domain of the CB2 Cannabinoid Receptor. *J Neurochem* 75(6):2485-2491.
- Rhee M-H *et al* (2000) Role of the highly conserved Asp-Arg-Tyr motif in signal transduction of the CB2 cannabinoid receptor. *FEBS Lett* 466(2-3):300-304
- Ribas C *et al* (2007) The G protein-coupled receptor kinase (GRK) interactome: Role of GRKs in GPCR regulation and signaling. *Biochim Biophys Acta* 1768(4):913-922
- Rinne A *et al* (2015) The mode of agonist binding to a G protein–coupled receptor switches the effect that voltage changes have on signaling. *Sci Signaling* 8(401):ra110
- Ritter SL, Hall RA (2009) Fine-tuning of GPCR activity by receptor-interacting proteins. *Nat Rev Mol Cell Biol* 10:819-830
- Ryan JA (2008) Evolution of Cell Culture Surfaces. *BioFiles* 3(8):21
- Röblitz S, Weber M (2013) Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Advances in Data Analysis and Classification* 7:147-179
- Schiffmann A, Gimpl G (2018) Sodium functions as a negative allosteric modulator of the oxytocin receptor. *BBA – Biomembranes* 1860:1301-1308
- Schmid N *et al* (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur Biophys J* 40:843-856

- Schneider EH *et al* (2010) Impact of the DRY Motif and the Missing “Ionic Lock” on Constitutive Activity and G-Protein Coupling of the Human Histamine H4 Receptor. *J Pharmacol Exp Ther* 333(2):383-392
- Schwantes CR, Pande VS (2013) Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J Chem Theory Comput* 9(4):2000-2009
- Schönege AM *et al* (2017) Evolutionary action and structural basis of the allosteric switch controlling β 2AR functional selectivity. *Nat Comm* 8:art. 2169
- Scott MG *et al* (2002) Differential nucleocytoplasmic shuttling of beta-arrestins. Characterization of a leucine-rich nuclear export signal in beta-arrestin2. *J Biol Chem* 277:37693-37701
- Sente A *et al* (2018) Molecular mechanism of modulating arrestin conformation by GPCR phosphorylation. *Nat Struct Mol Biol* 25:538-545
- Shalaeva DN *et al* (2015) Eukaryotic G protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins. *Biol Direct* 10(63):1-12
- Shaw DE *et al* (2008) Anton, a special-purpose machine for molecular dynamics simulation. *Commun ACM* 51(7):91-97
- Shaw DE *et al* (2009) Millisecond-scale molecular dynamics simulations on Anton. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, DOI: 10.1145/1654059.1654126, page consultée le 31 août 2019
- Shukla AK *et al* (2013) Structure of active β -arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. *Nature* 497:137-141
- Slack RJ, Hall DA (2012) Development of operational models of receptor activation including constitutive receptor activity and their use to determine the efficacy of the chemokine CCL17 at the CC chemokine receptor CCR4. *Br J Pharmacol* 166:1774-1792
- Sorzano COS *et al* (2014) A Survey of Dimensionality Reduction Techniques. arXiv, <https://arxiv.org/abs/1403.2877>, page consultée le 28 août 2019.
- Sounier R *et al* (2015) Propagation of conformational changes during μ -opioid receptor activation. *Nature* 524:375-378
- Spalding TA *et al* (1998) Identification of a Ligand-dependent Switch within a Muscarinic Receptor. *J Biol Chem* 273(34):21563-21568

Szokodi I *et al* (2002) Apelin, the Novel Endogenous Ligand of the Orphan Receptor APJ, Regulates Cardiac Contractility. *Circ Res* 91:434-440

Tang M *et al* (2003) Regulator of G-protein signaling-2 mediates vascular smooth muscle relaxation and blood pressure. *Nat Med* 9(12):1506-1512

Tao YX (2000) Constitutive Activation of G Protein-Coupled Receptors as a Result of Selective Substitution of a Conserved Leucine Residue in Transmembrane Helix III. *Mol Endocrinol* 14(8):1272-1282

Tao YX (2008) Constitutive activation of G protein-coupled receptors and diseases: Insights into mechanisms of activation and therapeutics. *Pharmacol Ther* 120(2):129-148

Tehan BG *et al* (2014) Unifying Family A GPCR Theories of Activation. *Pharmacol Ther* 143(1):51-60

Thompson AA *et al* (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485:395-399

Thompson RC (1972) Principal Submatrices. VIII. Principal Sections of a Pair of Forms. *Rocky Mountain Journal of Mathematics* 2(1):97-110

Thomsen ARB *et al* (2016) GPCR-G Protein- β -Arrestin Super-Complex Mediates Sustained G Protein Signaling. *Cell* 166(4):907-919

Tonacchera M *et al* (1998) Hyperfunctioning Thyroid Nodules in Toxic Multinodular Goiter Share Activating Thyrotropin Receptor Mutations with Solitary Toxic Adenoma. *J Clin Endocrinol Metab* 83(2):492-498

Trzakowski B *et al* (2012) Action of Molecular Switches in GPCRs - Theoretical and Experimental Studies. *Curr Med Chem* 288(4):2593-2604

Trần K *et al* (2018) A Systematic Exploration of Macrocyclization in Apelin-13: Impact on Binding, Signaling, Stability, and Cardiovascular Effects. *J Med Chem* 61:2266-2277

Vaisse C *et al* (2000) Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity. *J Clin Invest* 106(2):253-262

van der Worp HB *et al* (2010) Can animal models of disease reliably inform human studies. *PLoS Med* 7(3):e1000245

Venkatakrishnan AJ *et al* (2013) Molecular signatures of G-protein-coupled receptors. *Nature* 494:185-194

Vsevolod K *et al* (2014) Allosteric sodium: a key co-factor in class A GPCR signaling. *Trends Biochem Sci* 39(5):233-244

- Wacker D *et al* (2013) Structural features for functional selectivity at serotonin receptors. *Science* 340:615-619
- Wang ZQ, Tao YX (2011) Functional studies on twenty novel naturally occurring melanocortin-4 receptor mutations. *Biochim Biophys Acta* 1812(9):1190-1199
- Warne T *et al* (2008) Structure of a β 1-adrenergic G-protein-coupled receptor. *Nature* 454:486-491
- Webb B, Šali A (2014) Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinf* 47:5.6.1-5.6.32
- Weber JK, Pande VS (2011) Characterization and Rapid Sampling of Protein Folding Markov State Model Topologies. *J Chem Theory Comput* 7(10):3405-3411
- Wedegaertner PB (2012) G Protein Trafficking. *Subcell Biochem* 63:193-223
- Weinan E, Vanden-Eijnden E (2010) Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annu Rev Phys Chem* 61:391-420
- Wennberg CL *et al* (2012) Large Influence of Cholesterol on Solute Partitioning into Lipid Membranes. *JACS* 134:5351-5361
- Wescott MP *et al* (2016) Signal transmission through the CXC chemokine receptor 4 (CXCR4) transmembrane helices. *PNAS* 113(35):9928-9933
- Wettschureck N, Offermanns S (2005) Mammalian G Proteins and Their Cell Type Specific Functions. *Physiol Rev* 85(4):1159-1204
- Whistler JL *et al* (2002) Constitutive Activation and Endocytosis of the Complement Factor 5a Receptor: Evidence for Multiple Activated Conformations of a G Protein-Coupled Receptor. *Traffic* 3(12):866-877
- White KL *et al* (2018) Structural Connection between Activation Microswitch and Allosteric Sodium Site in GPCR Signaling. *Structure* 26:259-269
- Whorton MR, MacKinnon R (2013) X-ray structure of the mammalian GIRK2- $\beta\gamma$ G-protein complex. *Nature* 498:190-197
- Wit E *et al* (2012) ‘All models are wrong...’: an introduction to model uncertainty. *Statistica Neerlandica* 66(3):217-236
- Wong SFK *et al* (2003) G Protein Selectivity Is Regulated by Multiple Intracellular Regions of GPCRs. *Neurosignals* 12:1-12

Xie G, Palmer PP (2007) How Regulators of G Protein Signaling Achieve Selective Regulation. *J Mol Biol* 366(2):349-365

Ye L *et al* (2016) Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature* 533:265-268

Ye L *et al* (2018) Mechanistic insights into allosteric regulation of the A2A adenosine G protein-coupled receptor by physiological cations. *Nat Comm* 9:art. 1372

Zhang Y *et al* (2017) Cryo-EM structure of the activated GLP-1 receptor in complex with a G protein. *Nature* 546:248-253

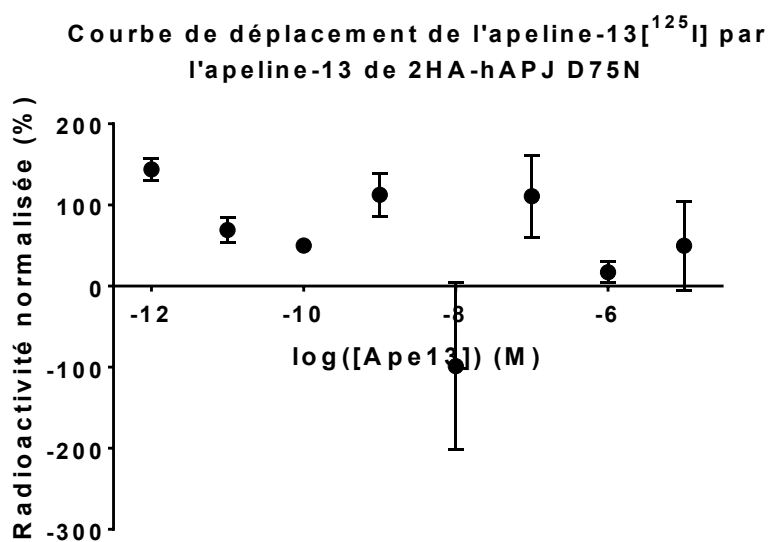
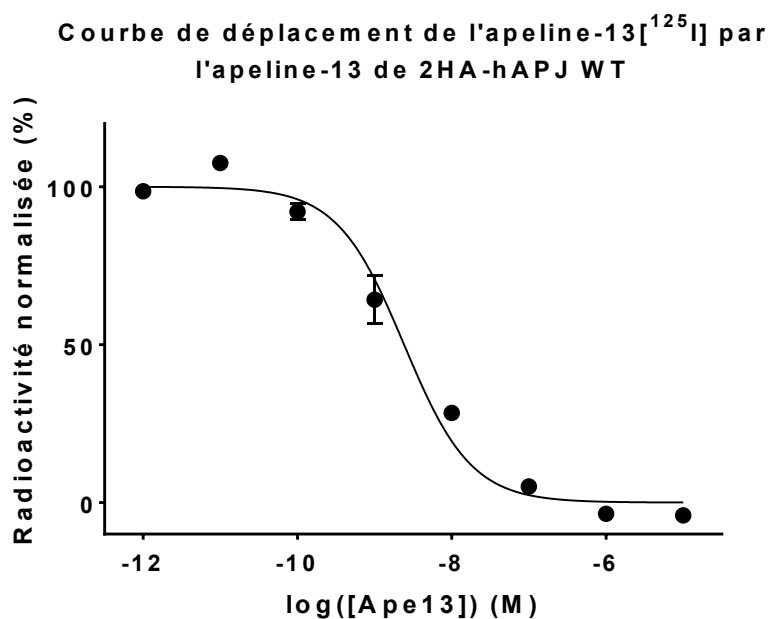
Zhou B *et al* (2019) Can Adding Constitutive Receptor Activity Redefine Biased Signaling Quantification. *Trends Pharmacol Sci* 40(3):156-160

Zhou XE *et al* (2016) X-ray laser diffraction for structure determination of the rhodopsin-arrestin complex. *Sci Data* 3:160021

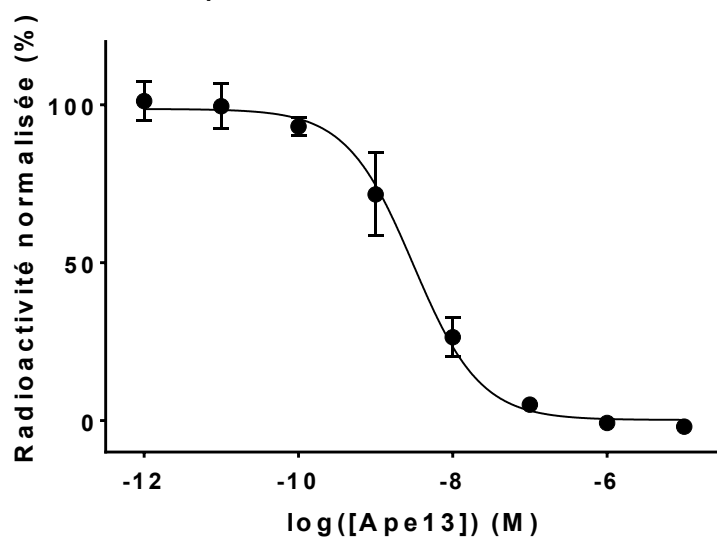
Zhou XE *et al* (2017) Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors. *Cell* 170:457-469e13

Annexe I : Courbes de déplacement de l'apeline radiomarquée par de l'apeline non-radiomarquée.

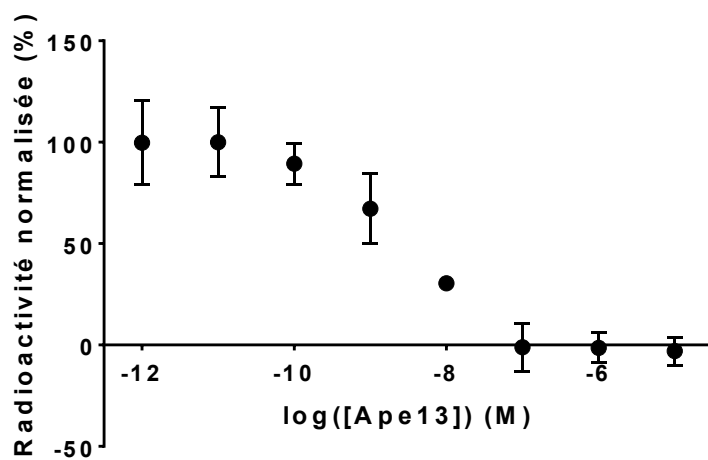
Les points et les barres d'erreurs représentent la moyenne \pm SEM de trois expériences indépendantes, effectuées chacune en triplicata.



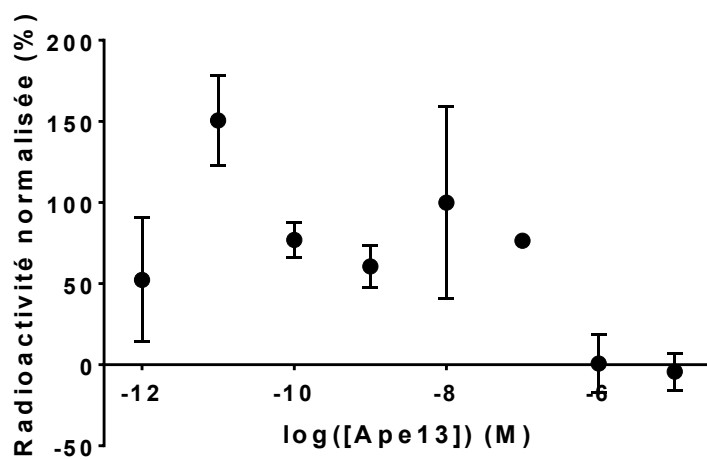
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ N112G



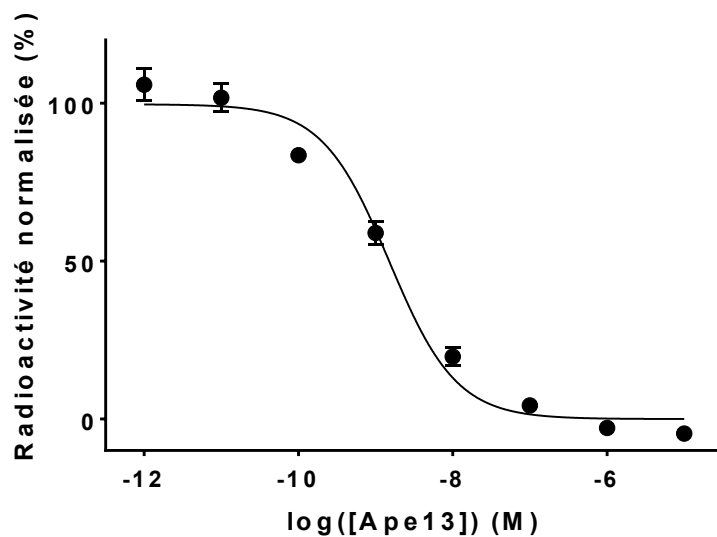
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ N112F



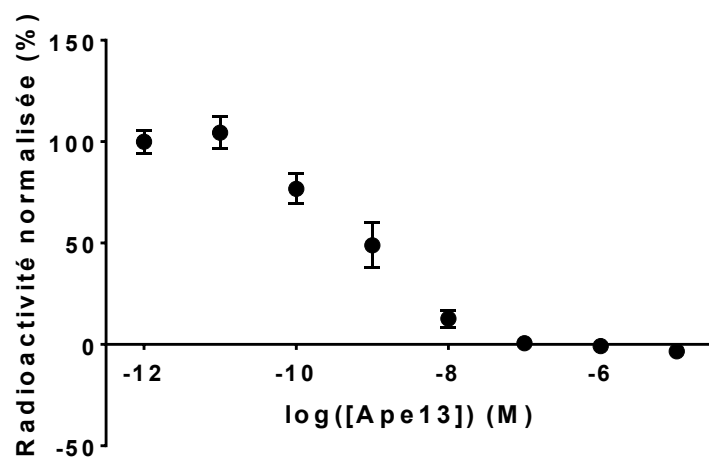
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ N112W



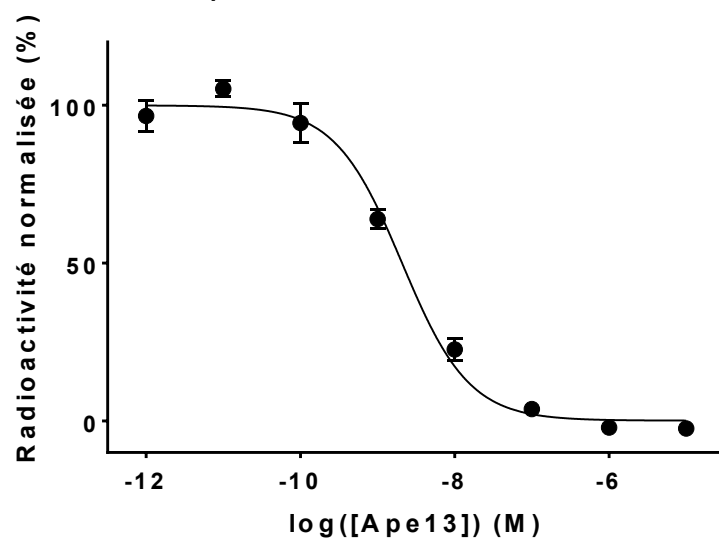
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ V117L



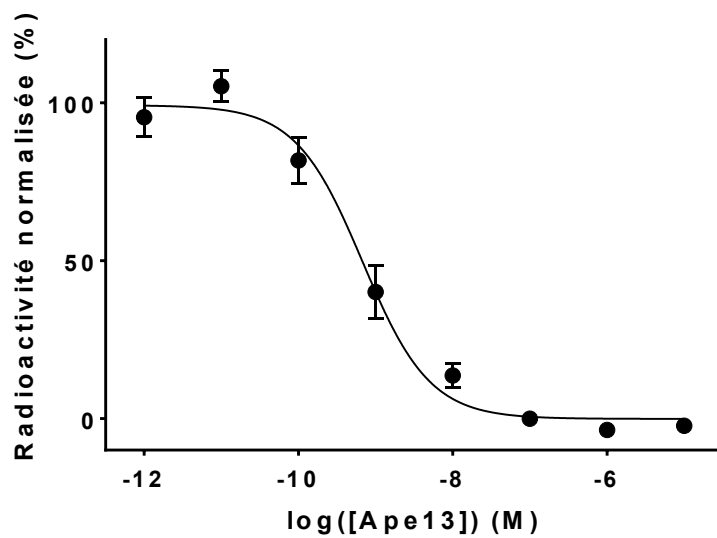
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ L120R



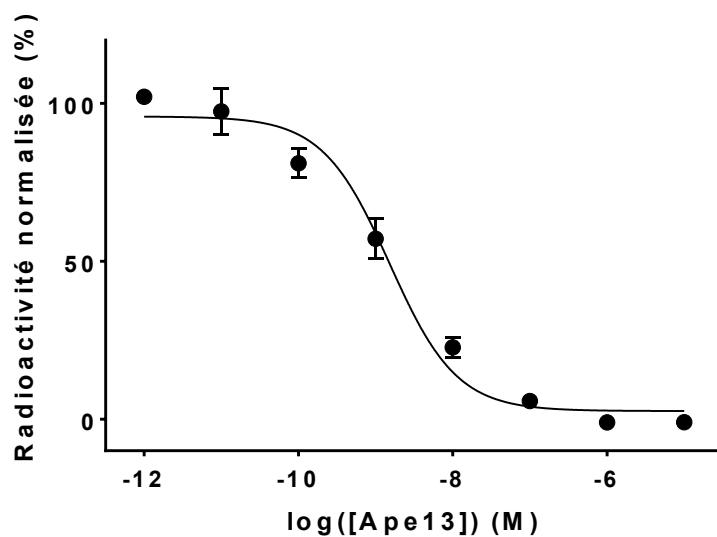
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ L120K



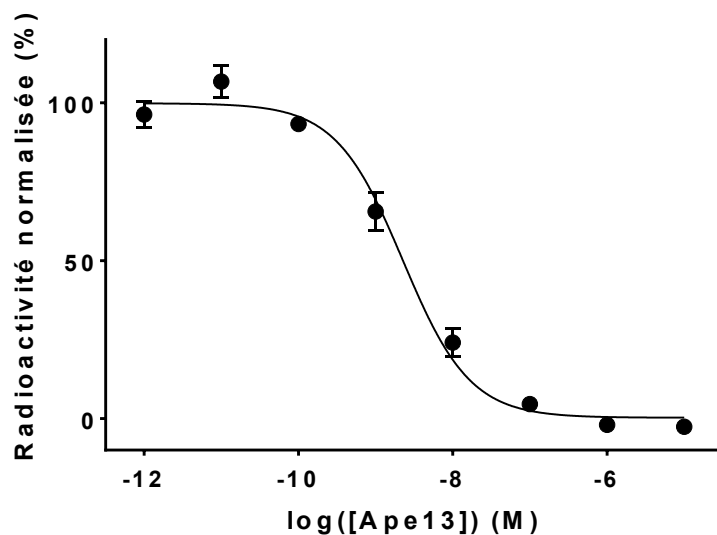
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ L120A



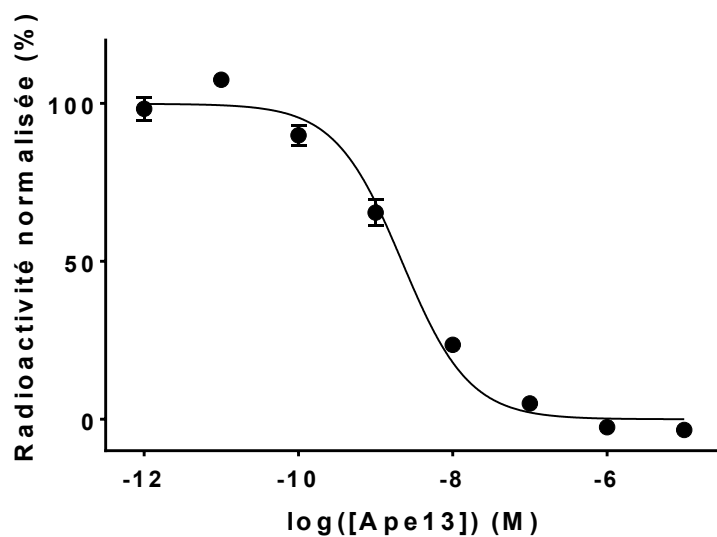
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ D126N



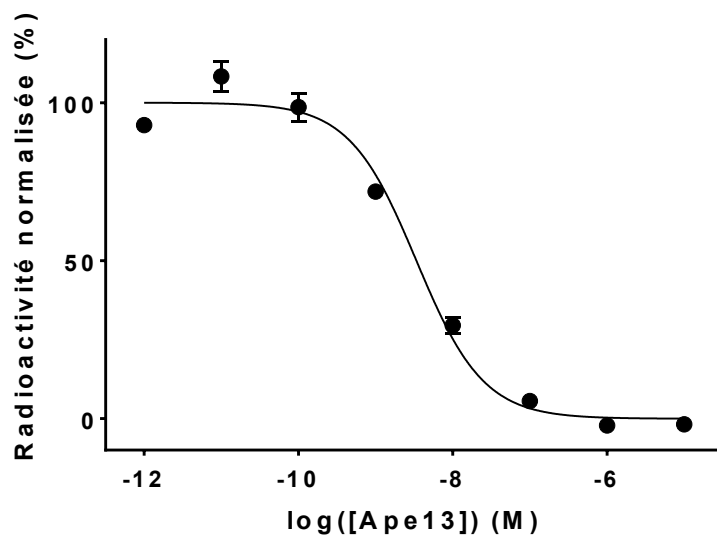
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ D126V



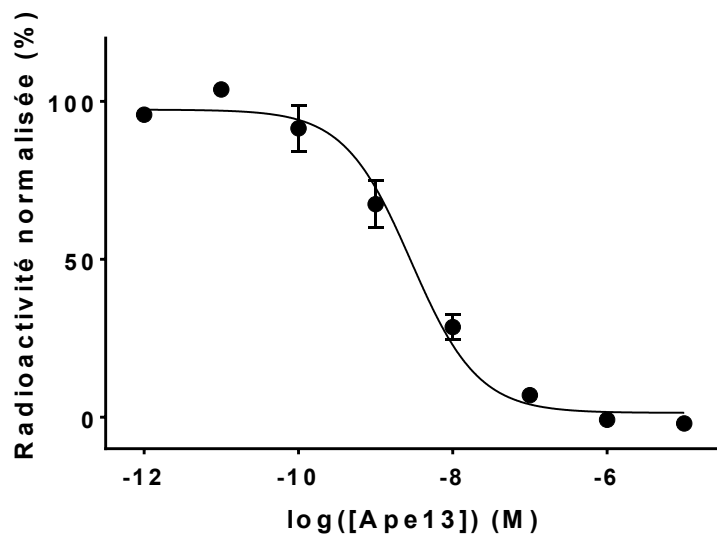
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ L253F



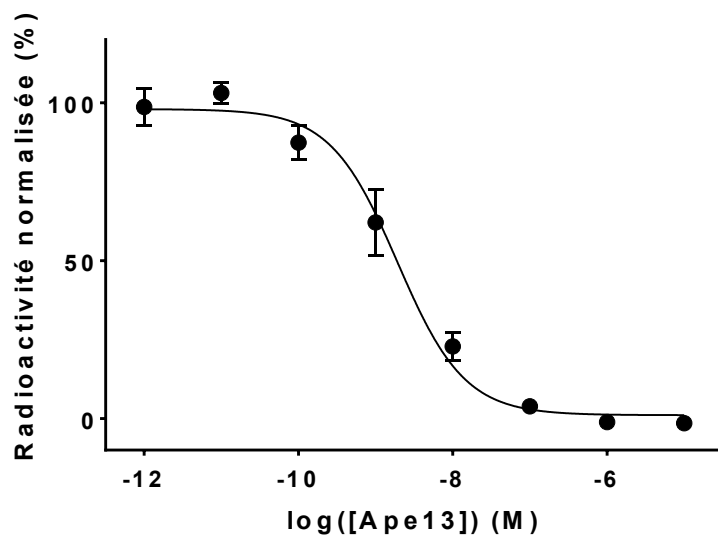
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ L253Q



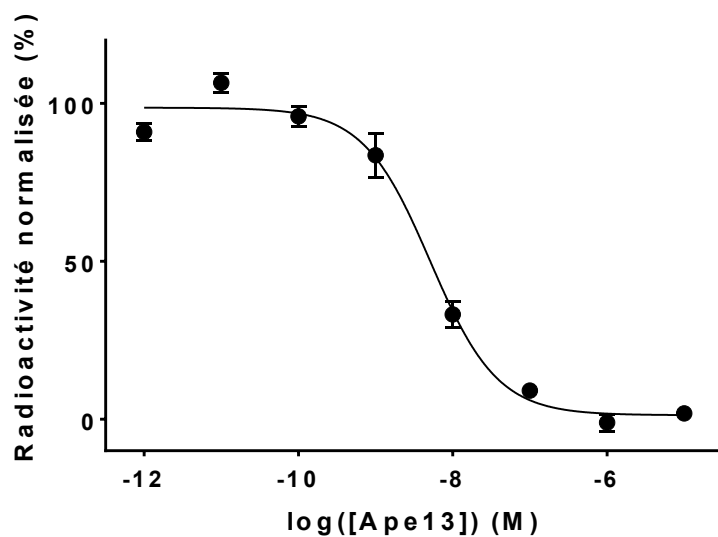
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ L253I



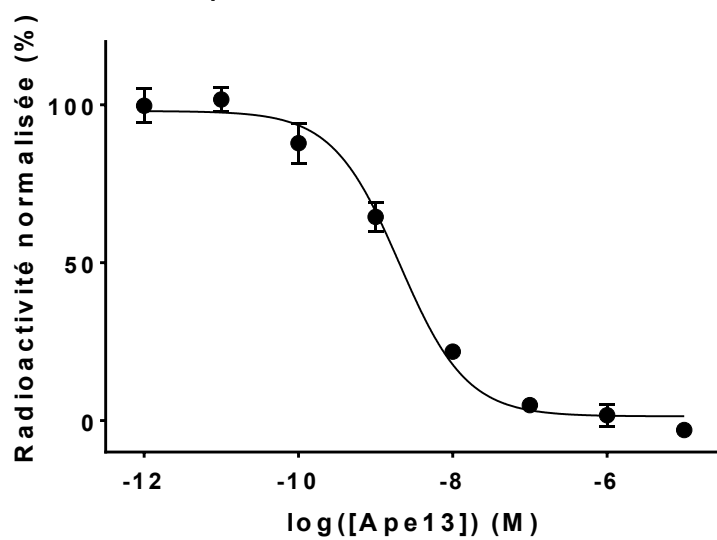
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ V254L



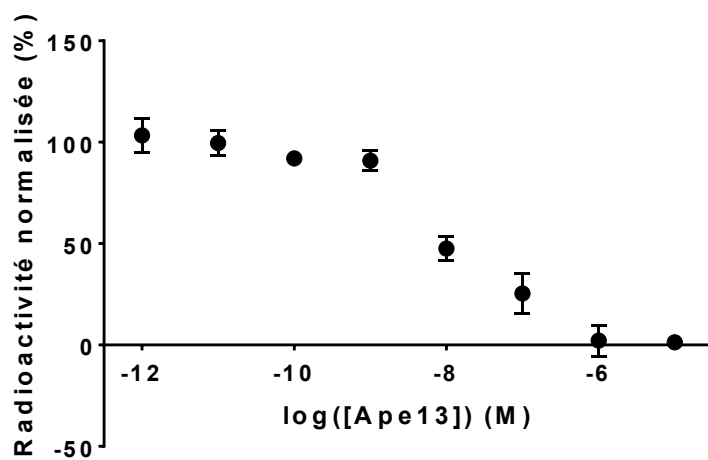
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ F257T



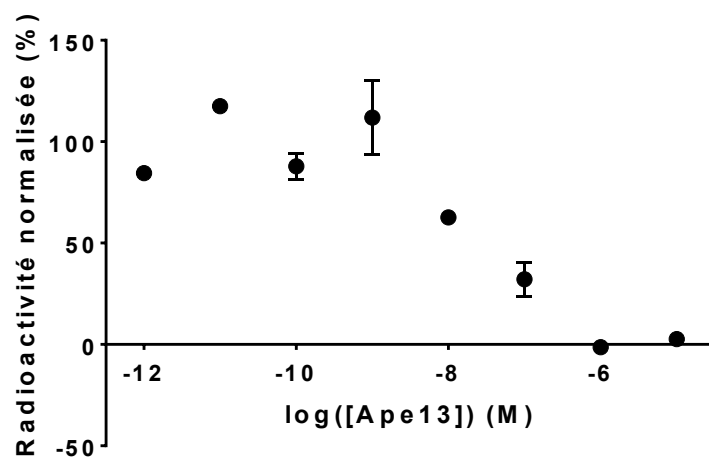
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ F257L



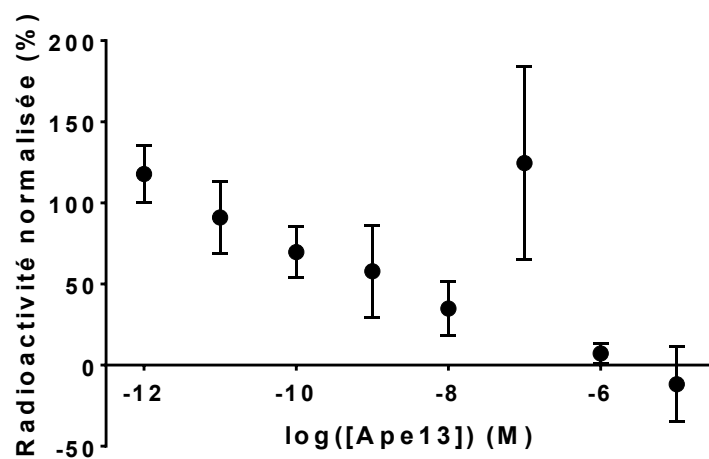
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ F257G



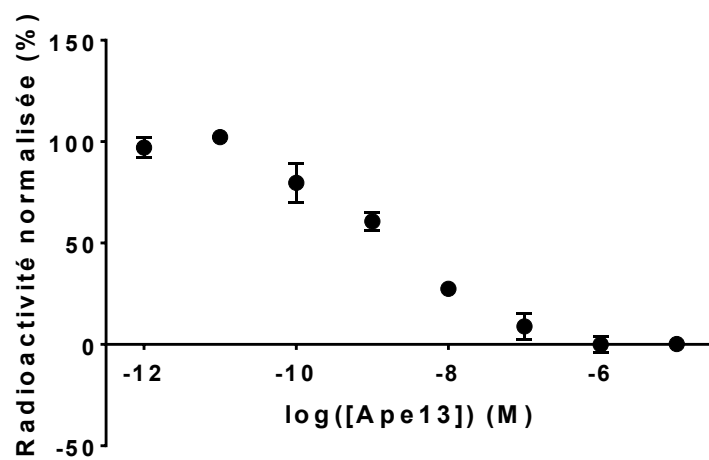
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ F257W



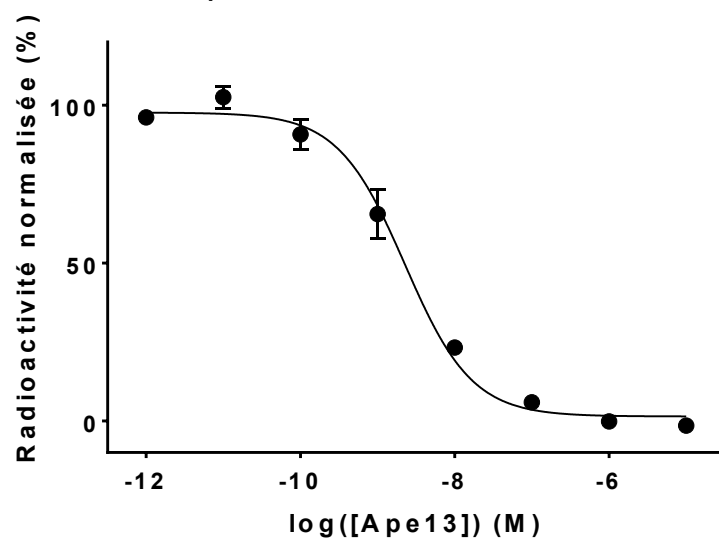
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ W261R

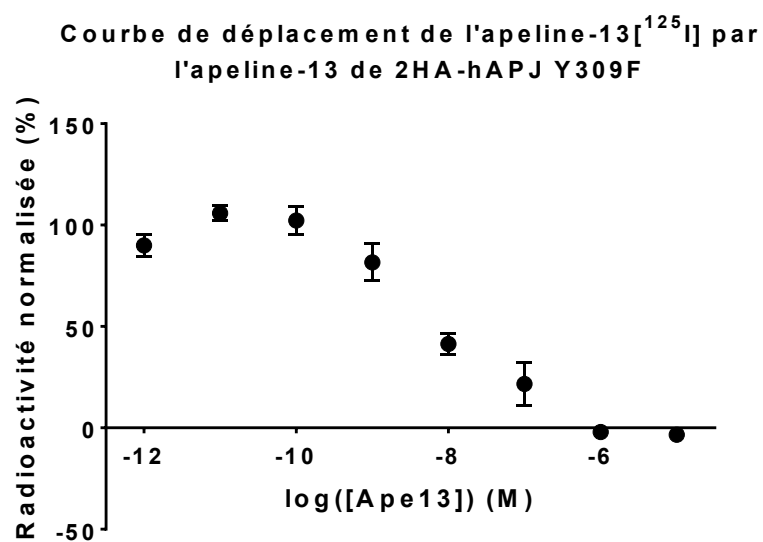
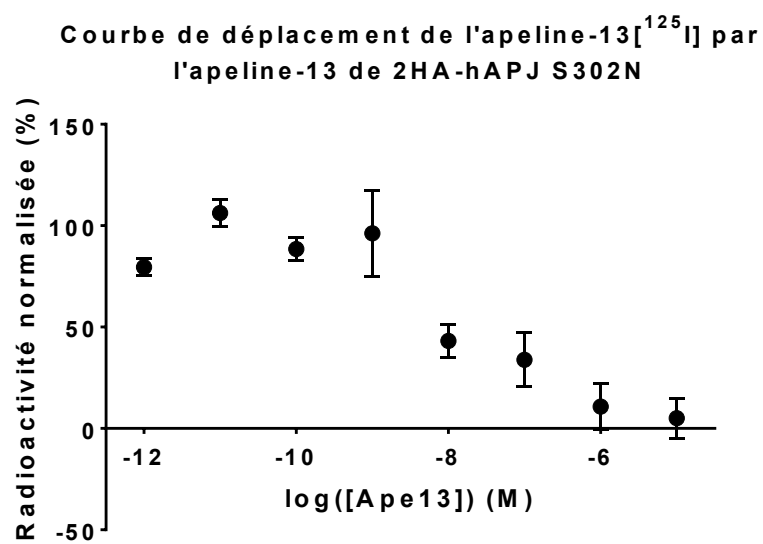


Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ W261F



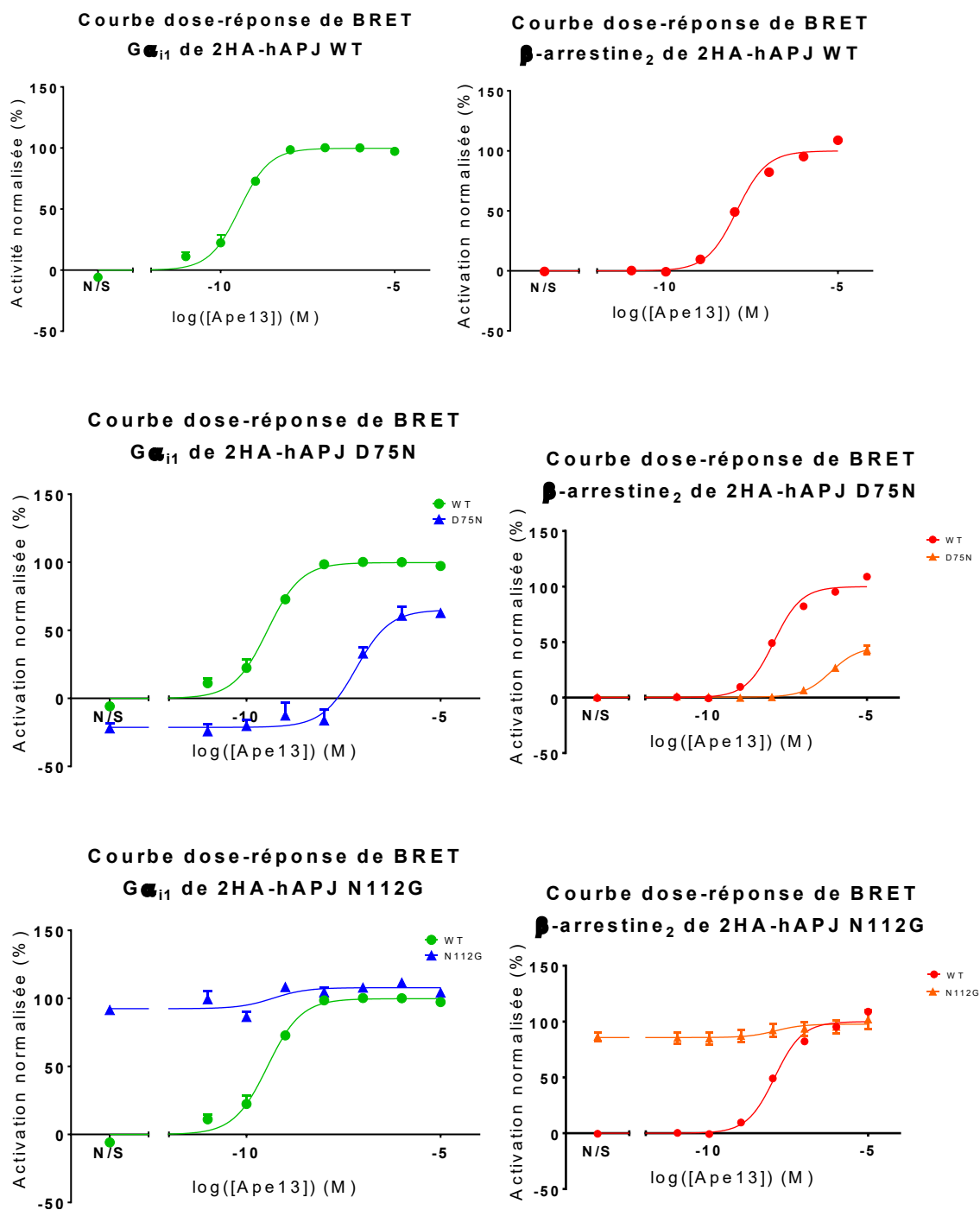
Courbe de déplacement de l'apeline-13 [125 I] par
l'apeline-13 de 2HA-hAPJ N301G





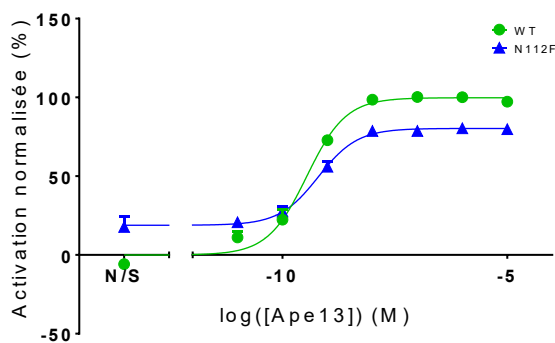
Annexe II : Courbes dose-réponse de BRET $G\alpha_{i1}$ et β -arrestine 2 pour l'ensemble des mutants étudiés.

L'ensemble des points représente la moyenne \pm SEM de trois expériences indépendantes.



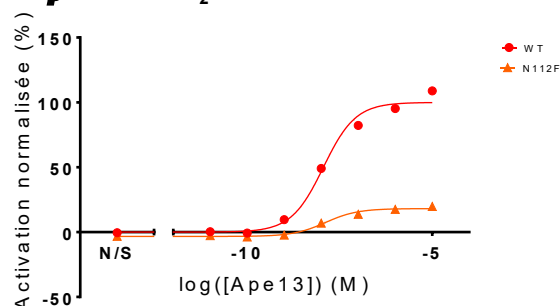
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ N112F



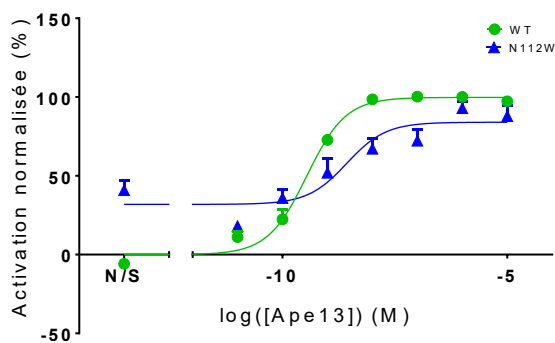
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ N112F



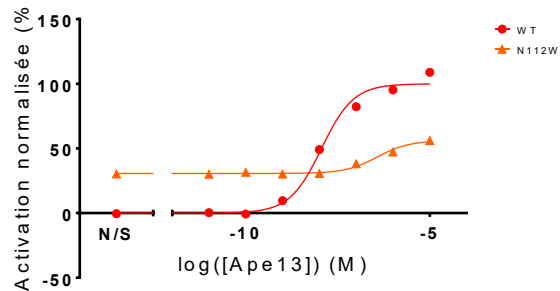
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ N112W



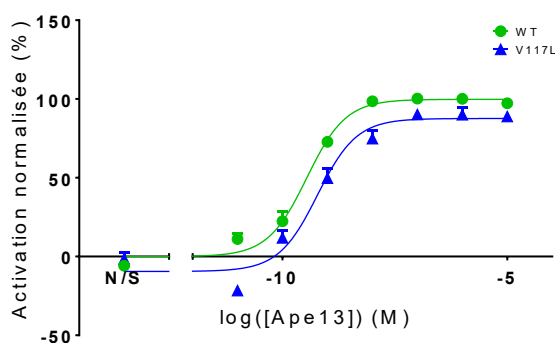
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ N112W



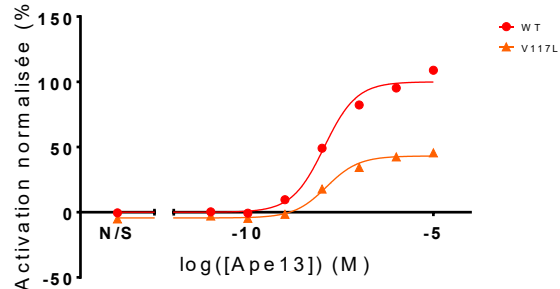
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ V117L

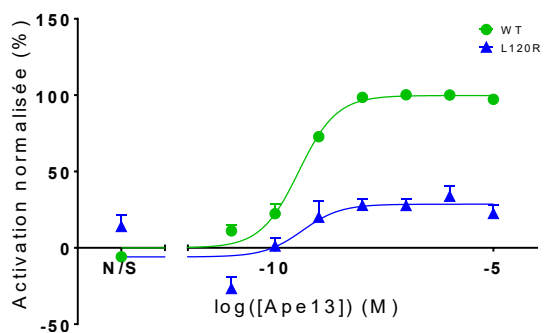


Courbe dose-réponse de BRET

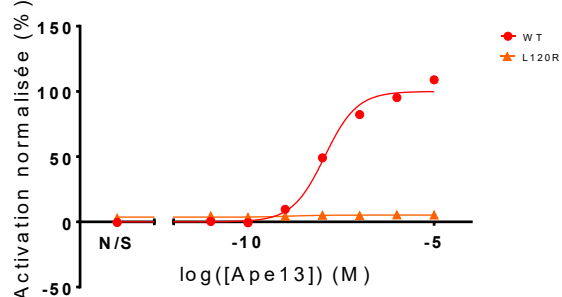
β -arrestine₂ de 2HA-hAPJ V117L



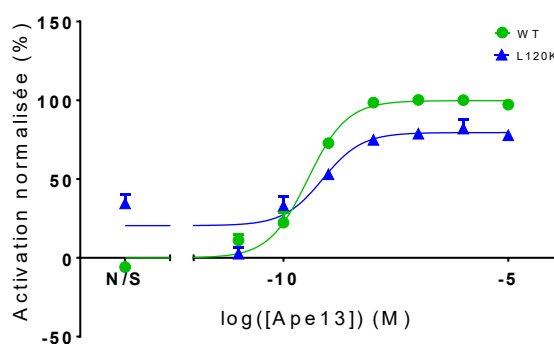
Courbe dose-réponse de BRET
G α_{i1} de 2HA-hAPJ L120R



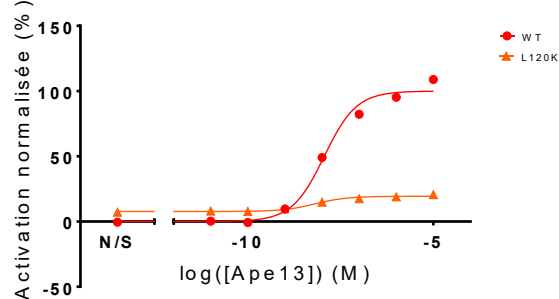
Courbe dose-réponse de BRET
 β -arrestine₂ de 2HA-hAPJ L120R



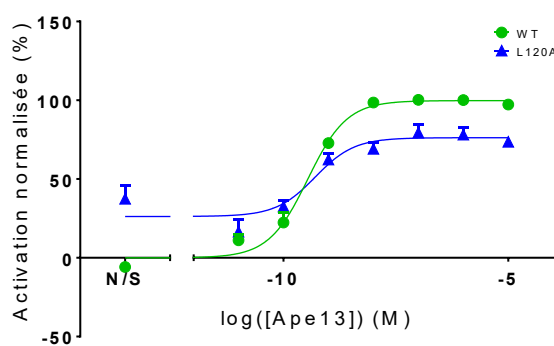
Courbe dose-réponse de BRET
G α_{i1} de 2HA-hAPJ L120K



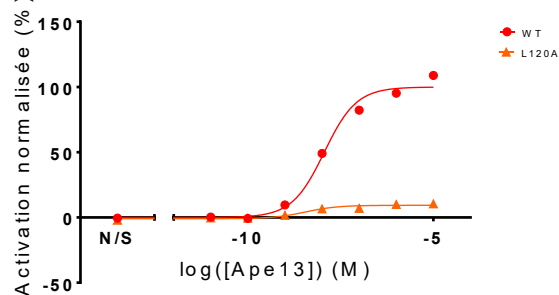
Courbe dose-réponse de BRET
 β -arrestine₂ de 2HA-hAPJ L120K



Courbe dose-réponse de BRET
G α_{i1} de 2HA-hAPJ L120A

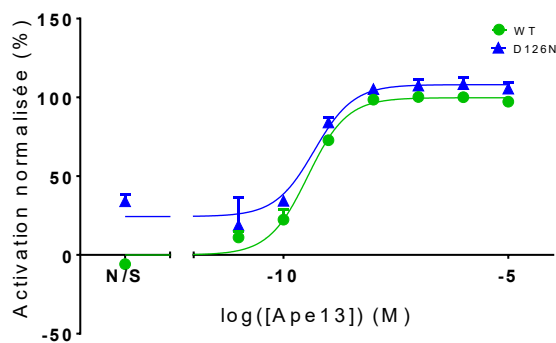


Courbe dose-réponse de BRET
 β -arrestine₂ de 2HA-hAPJ L120A



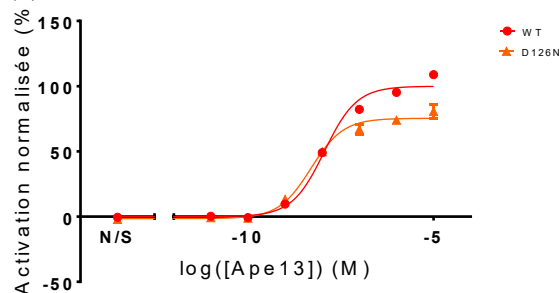
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ D126N



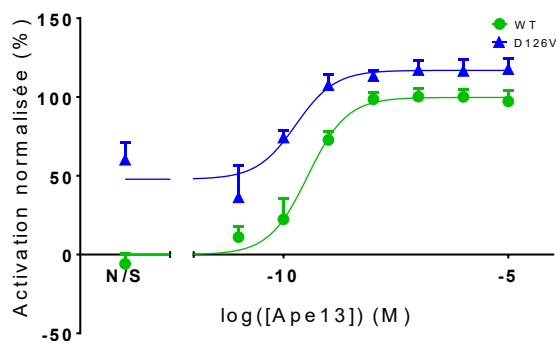
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ D126N



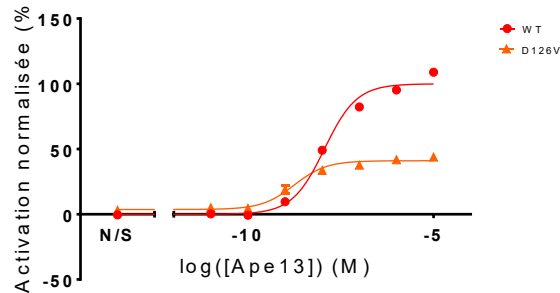
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ D126V



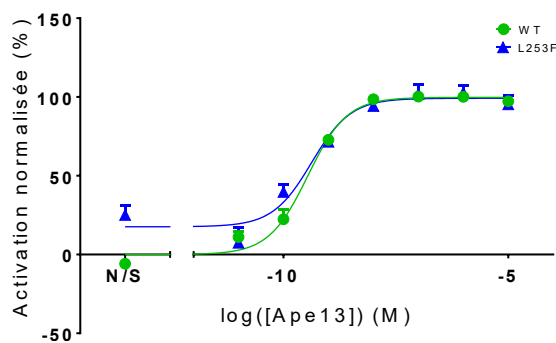
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ D126V



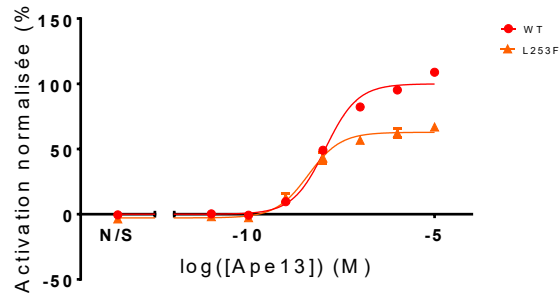
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ L253F



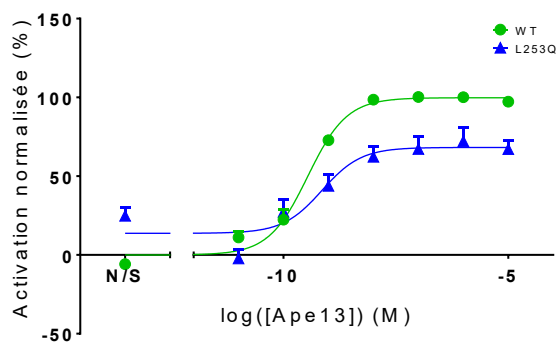
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ L253F



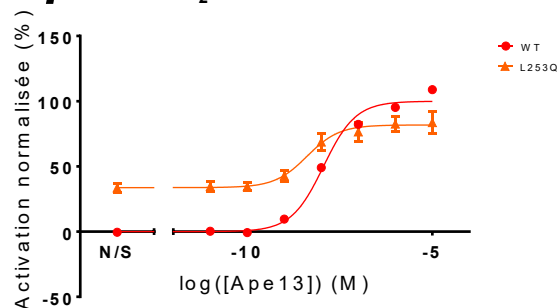
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ L253Q



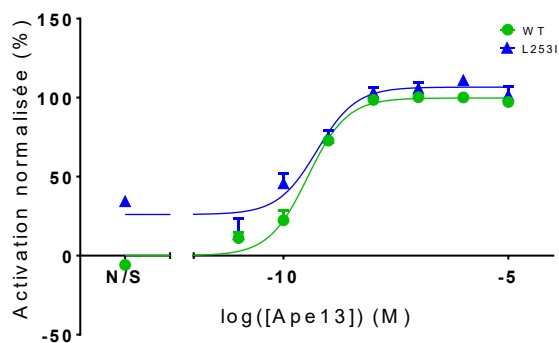
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ L253Q



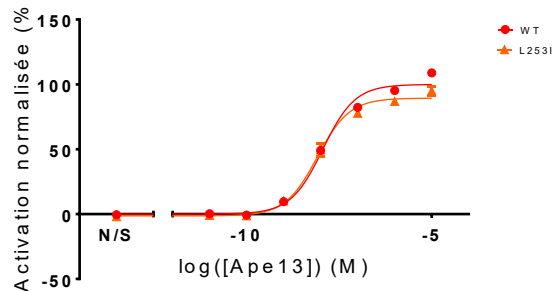
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ L253I



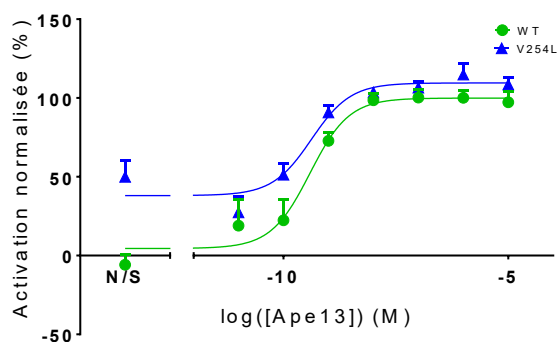
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ L253I



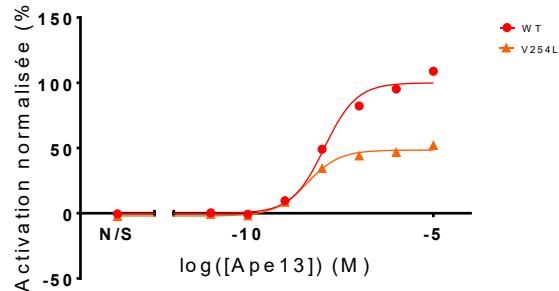
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ V254L



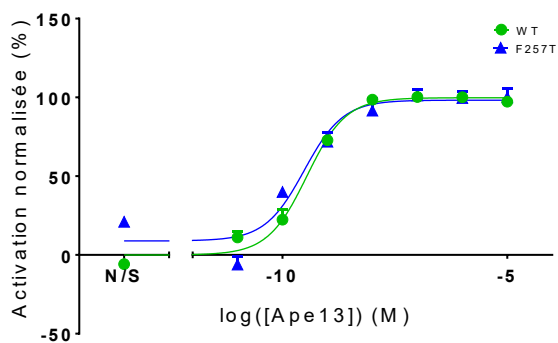
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ V254L



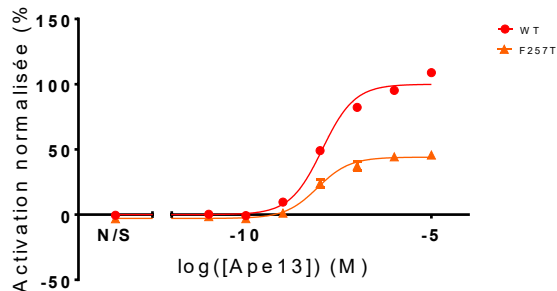
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ F257T



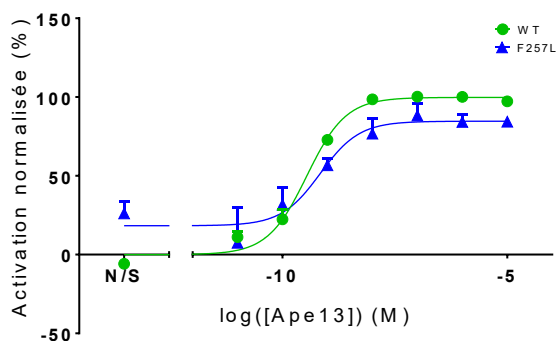
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ F257T



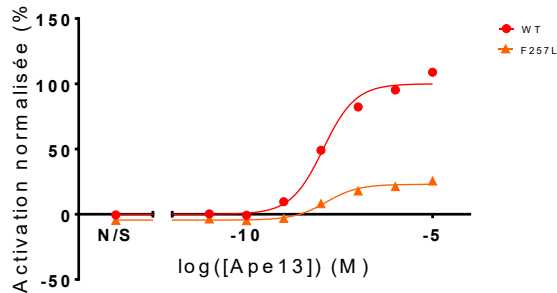
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ F257L



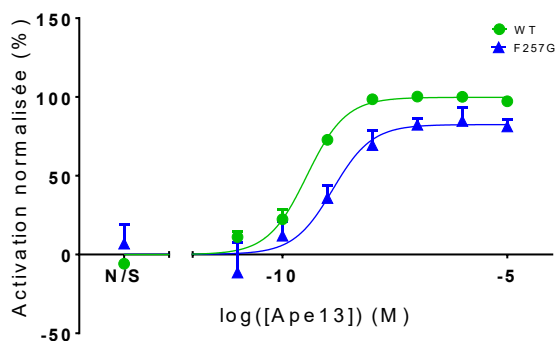
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ F257L



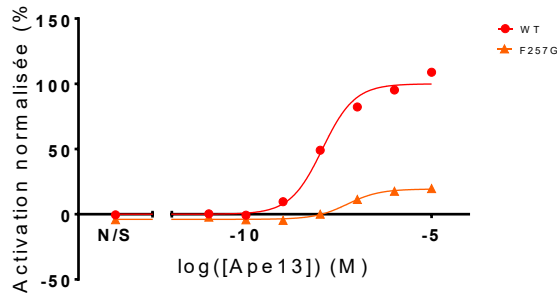
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ F257G



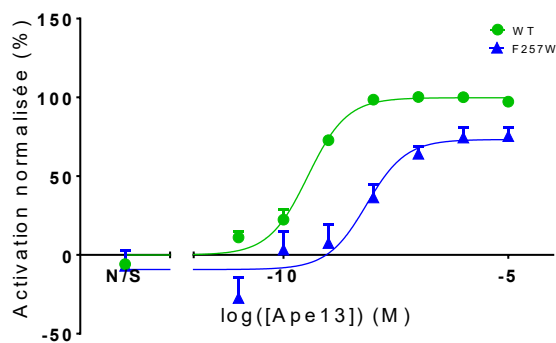
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ F257G



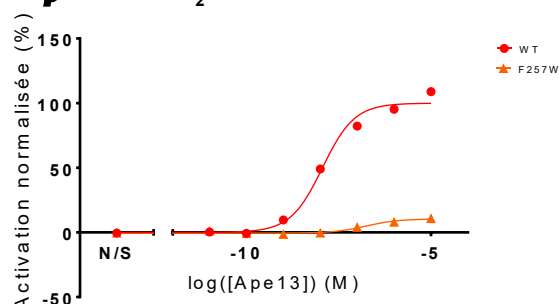
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ F257W



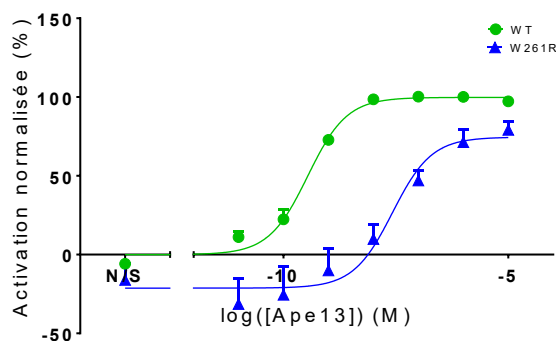
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ F257W



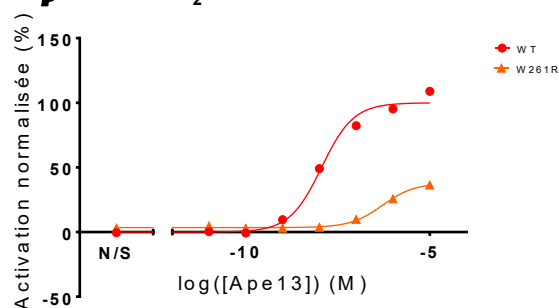
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ W261R



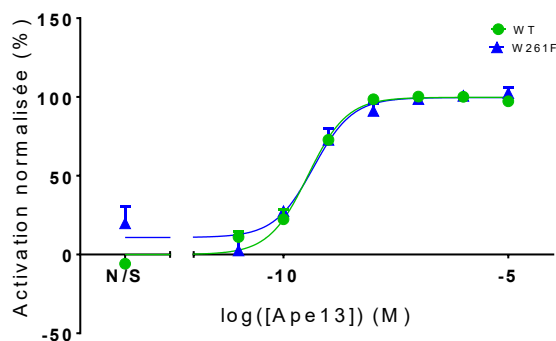
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ W261R



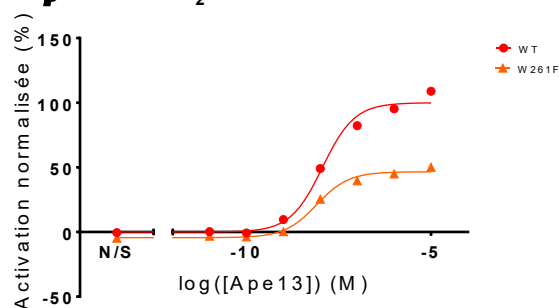
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ W261F



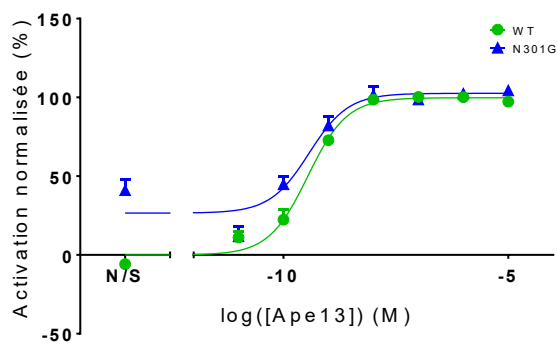
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ W261F



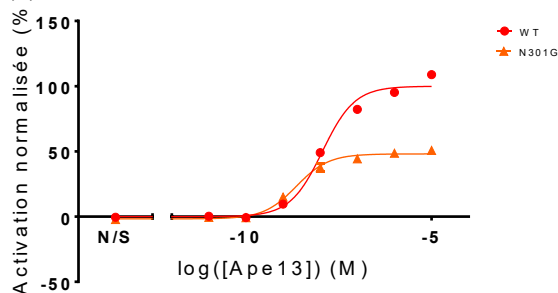
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ N301G



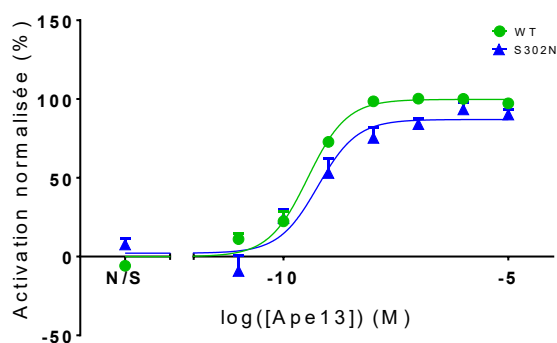
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ N301G



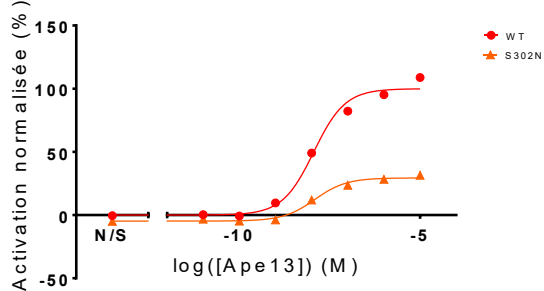
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ S302N



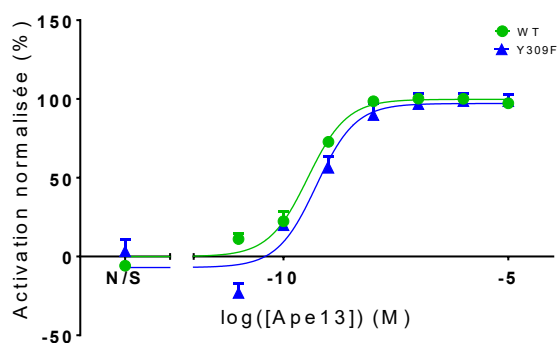
Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ S302N



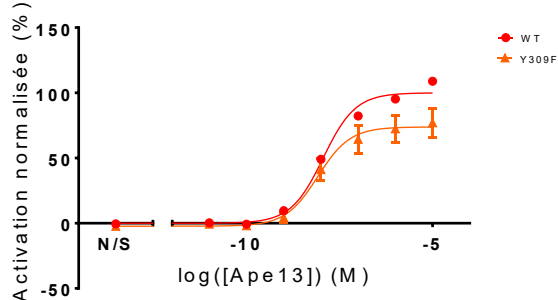
Courbe dose-réponse de BRET

G α_{i1} de 2HA-hAPJ Y309F



Courbe dose-réponse de BRET

β -arrestine₂ de 2HA-hAPJ Y309F



Annexe III : Inventaire des mutants d'intérêt répertoriés dans la littérature sur les GPCRs

Inventaire (probablement incomplet) des différents mutants répertoriés dans la littérature sur les GPCRs. Légende : CA : Activité constitutive AM : Activité maximale IN : Inactif (parfois pairé avec NB) NB : Absence de liaison du ligand REC50 : Ratio d'EC50 (EC50(mutant)/EC50(WT)) DBA : Diminution de l'activité basale AGI : Activité résiduelle après exposition à un agoniste inverse					
Récepteur	Région	Mutation (Notation Ballesteros- Weinstein)	Effet	Valeur	Source et remarques
AT1	TM3	WT	CA	5	(Feng Y <i>et al</i> , 1998)
		N3.35Y	CA	<5	
		N3.35F	CA	<5	
		N3.35H	CA	<5	
		N3.35K	CA	<5	
		N3.35Q	CA	<5	
		N3.35I	CA	<5	
		N3.35C	CA	10	
		N3.35A	CA	10	
		N3.35S	CA	15	
		N3.35G	CA	50	
		WT	AM	100	
		N3.35Y	AM	70	
		N3.35F	AM	45	
		N3.35H	AM	50	
		N3.35K	AM	35	
		N3.35Q	AM	55	
		N3.35I	AM	40	
		N3.35C	AM	80	
		N3.35A	AM	90	
		N3.35S	AM	100	
		N3.35G	AM	100	
M5 Muscarinique	TM6	WT	CA	10	(Spalding TA <i>et al</i> , 1998) Pas de données de binding
		F6.44G	CA	10	

		F6.44A	CA	40	Observation de la croissance cellulaire
		F6.44V	CA	90	
		F6.44T	CA	90	
		F6.44L	CA	60	
		F6.44W	CA	0	
		WT	AM	100	
		F6.44G	AM	120	
		F6.44A	AM	200	
		F6.44V	AM	180	
		F6.44T	AM	240	
		F6.44L	AM	160	
		F6.44W	AM	60	
		F6.44G	REC50	5000	
		N6.52H	CA	50	
TSHR	TM6	WT	CA	20	(Tonacchera M <i>et al</i> , 1998) Pas de données de binding. Observation du comportement tumorigène.
		L6.40F	CA	90	
		I6.41L	CA	55	
		P6.50S	CA	100	
B2-R		WT	CA	100	(Marie J <i>et al</i> , 1999) N3.35A et W6.48F transforment un agoniste inverse en agoniste
	TM2	D2.50N	CA	100	
	TM3	N3.35A	CA	1850	
	TM6	W6.48F	CA	400	
		W6.48Q	CA	350	
	TM7	Y7.43F	CA	100	
		WT	AM	100	
	TM2	D2.50N	AM	45	
	TM3	N3.35A	AM	367	
	TM6	W6.48F	AM	200	
		W6.48Q	AM	170	
	TM7	Y7.43F	AM	100	
C5aR		WT	CA	0	(Baranski TJ <i>et al</i> , 1999) Pas de données de binding. Observation de la croissance de levures exprimant C5aR ± C5a
	TM3	I3.40N	CA	+	
		L3.43Q	CA	+	
	TM6	F6.44A	CA	++	
		WT	AM	+++	
	TM3	T3.45A	AM	+	
		I3.48A	AM	+	
	TM5	L5.43A	AM	+	
		L5.43I	AM	+	
	TM6	V6.53A	AM	+	
		S6.59A	AM	+	

	TM3	L3.28A	IN/NB	----	
		L3.42A	IN/NB	----	
	TM5	A5.39V	IN/NB	----	
		R5.42H	IN/NB	----	
		Y5.58A	IN/NB	----	
	TM7/H8	D7.35A	IN/NB	----	
		V7.39A	IN/NB	----	
		V7.39L	IN/NB	----	
		S7.40A	IN/NB	----	
		N7.49A	IN/NB	----	
CXCR2		WT	CA	100	(Burger M <i>et al</i> , 1999)
	TM3	D3.49V	CA	150	
LHR	TM1	A1.46V	CA	?	(Fanelli F, 2000) Revue de littérature (pas de résultats)
	TM2	M2.43T	CA	?	
	TM3	L3.43R	CA	?	
	TM5	I5.54L	CA	?	
	TM6	D5.76G	CA	?	
		A6.34V	CA	?	
		M6.37I	CA	?	
		A6.38V	CA	?	
		I6.41L	CA	?	
		T6.43I	CA	?	
		D6.44G	CA	?	
		D6.44H	CA	?	
		C6.47R	CA	?	
	TM5	C5.55R	IN	----	
	TM7/H8	S7.46Y	IN	----	
		I7.55K	IN	----	
MC4R	Nterm	T11S	CA	?	(Vaisse C <i>et al</i> , 2000) Étude de patients obèses morbides
		R18C	CA	?	
	TM3	T3.53I	CA	?	
	TM4	R4.41W	CA	?	
	TM6	L6.40Q	CA	?	
	TM7/H8	I7.52T	CA	?	
FSHR	TM3	WT	CA	1	(Tao YX, 2000) Comportement différent de 3.43 chez récepteurs à haute activité basale (e.g. β 2-AR)
		L3.43R	CA	5	
		WT	AM	100	
		L3.43R	AM	25	
		L3.43K	AM	20	
		L3.43A	AM	50	
		L3.43D	AM	25	

β 2-AR	TM3	WT	CA	1	
		L3.43R	CA	8	
		L3.43K	CA	8	
		L3.43A	CA	18	
CB2R	WT		AM (Gi)	100	(Rhee M-H <i>et al</i> , 2000)
	TM3	D3.49A	AM (Gi)	NB	
		R3.50A	AM (Gi)	50	
		Y3.51A	AM (Gi)	40	
AT1		WT	CA	100	(Parnot C <i>et al</i> , 2000) Mutagenèse par saturation
	TM2	F2.53Y	CA	111	
	TM3	N3.35S	CA	170	
		L3.36H	CA	141	
		L3.42H	CA	116	
	TM5	L5.39P	CA	118	
	TM6	I6.40T	CA	111	
	TM7/H ₈	L7.56Q	CA	143	
FSHR	WT		CA (AMPc)	100	(Tao YX <i>et al</i> , 2000) Premier avec profil fonctionnel plus spécifique
	TM3	L3.43R	CA (AMPc)	500	
		L3.43K	CA (AMPc)	75	
		L3.43A	CA (AMPc)	100	
		L3.43D	CA (AMPc)	125	
	WT		AM	100	
	TM3	L3.43R	AM	25	
		L3.43K	AM	20	
		L3.43A	AM	60	
		L3.43D	AM	25	
	WT		CA (Internalisation)	0.15	
	TM3	L3.43R	CA (Internalisation)	0.45	
		L3.43K	CA (Internalisation)	0.6	
		L3.43A	CA (Internalisation)	0.15	
		L3.43D	CA (Internalisation)	0.05	

			tion)		
β 2-AR	WT		CA (AMPc)	1	
	TM3	L3.43R	CA (AMPc)	9	
		L3.43K	CA (AMPc)	9	
		L3.43A	CA (AMPc)	18	
	WT		AM	1	
	TM3	L3.43R	AM	1.25	
		L3.43K	AM	1	
		L3.43A	AM	1.5	
β 2-AR	WT		CA	1	(Ballesteros JA <i>et al</i> , 2001) Appui de l'existence du verrou ionique
	TM3	D3.49N	CA	>100	
	TM6	E6.30A	CA	>50	
		E6.30Q	CA	>10	
A3 AR	WT		CA	100	(Chen A <i>et al</i> , 2001)
	TM1	N1.50A	CA	98	
	TM3	R3.50K	CA	130	
		R3.50A	CA	150	
		R3.50N,E,H,F	CA	100	
	TM6	A6.34E	CA	160	
	WT		AM	502	
	TM1	N1.50A	AM	375	
	TM3	D3.49N,K,R	AM	450	
		R3.50K,A,E,H	AM	425	
		R3.50N	AM	387	
		Y3.51F	AM	342	
	TM6	A6.34E	AM	413	
	TM7	Y7.53F	AM	287	
	WT		REC50	100	
	TM3	Y3.51F	REC50	1200	
		A3.34E	REC50	26	
	TM7	Y7.53F	REC50	2700	
μ -OR	WT		CA	100	(Li J <i>et al</i> , 2001) D3.49E : agoniste inverse devient agoniste
	TM3	D3.49H	CA	250	
		D3.49Y	CA	250	
		D3.49Q	CA	225	
		D3.49M	CA	225	
		D3.49E	CA	50	
α 1-AR	WT		CA	14	(Greasley PJ <i>et al</i> , 2002)
	TM6	F6.27A	CA	59	

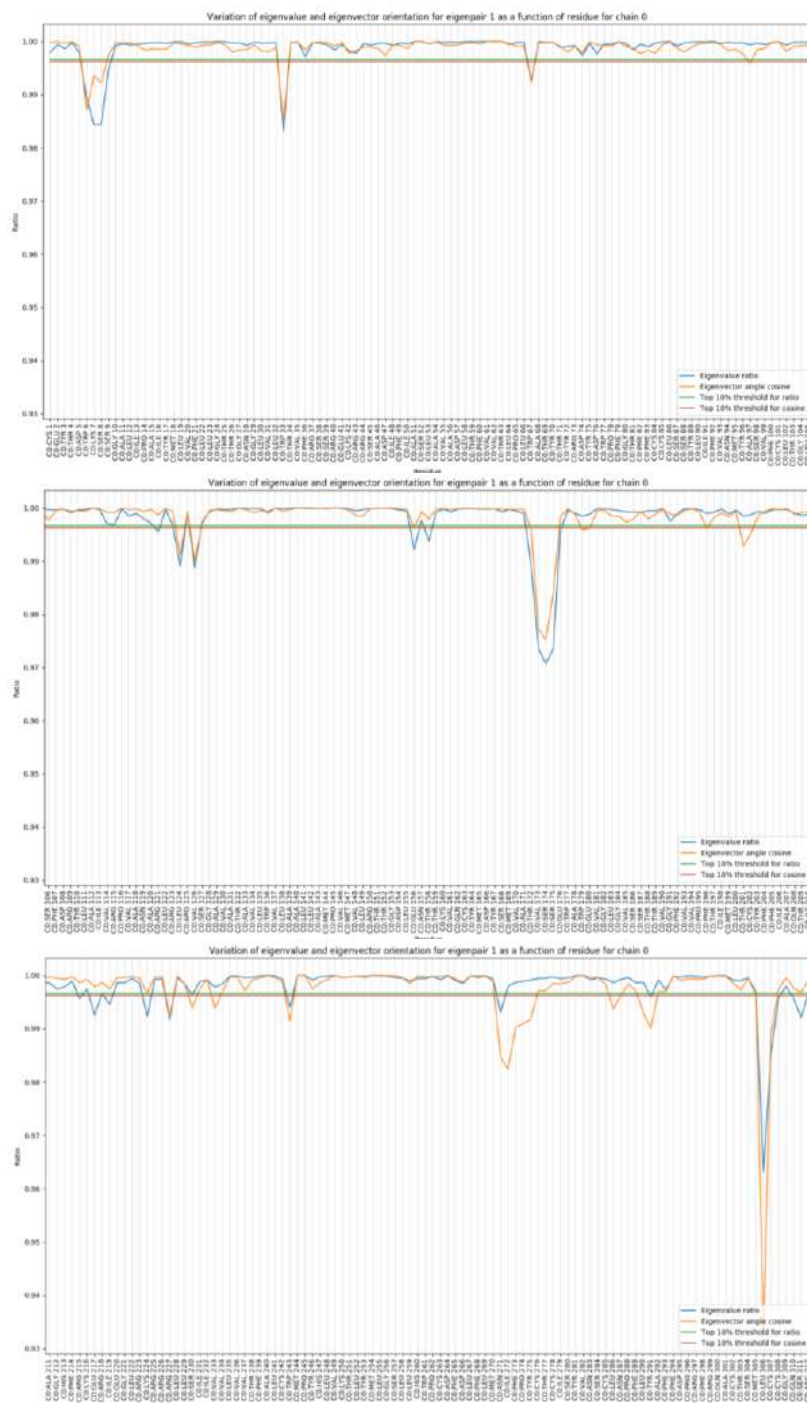
		E6.30A	CA	426	
		E6.30F	CA	252	
		E6.30D	CA	167	
		E6.30R	CA	301	
		E6.30K	CA	1386	
		F6.44L	CA	225	
		WT	AM	313-432	
	TM6	F6.27E	AM	291	
		A6.33E	AM	89	
		L6.37A	AM	106	
		L6.37F	AM	120	
		F6.44G	AM	8	
		F6.44A	AM	48	
		F6.44Y	AM	227	
C5aR	WT		CA	0	(Whistler JL <i>et al</i> , 2002)
	TM3	I3.40N	CA	+	
		L3.43Q	CA	+	
		I3.40N/L3.43Q	CA	++++	
	TM6	F6.44A	CA	+++	
	TM7	N7.49A	IN/NB		
	TM3/TM7	I3.40N/L3.43Q/N7.49A	CA	0	
	WT		AM	++++	
	TM3	I3.40N	AM	+++++	
		L3.43Q	AM	+++++	
		I3.40N/L3.43Q	AM	+++++	
	TM6	F6.44A	AM	++++	
	TM3/TM7	I3.40N/L3.43Q/N7.49A	AM	++++	
δ -OR	WT		CA	100	(Décaillot FM <i>et al</i> , 2003)
	TM3	D3.32N	CA	150	
	TM1	R1.24Q	CA	150	
	ICL2	F159L	CA	175	
	TM4	N4.46S	CA	175	
		I4.47T	CA	200	
	TM5	T5.38S	CA	200	
		K5.39R	CA	200	
	TM6	M6.36T	CA	150	
		W6.48R	CA	175	
	TM7/H8	Y7.43H	CA	250	
		L7.56H	CA	150	

		E7.58K	CA	150	
		C7.63R	CA	250	
		K7.70M	CA	150	
CB2R	WT		AM (Gi)	100	(Feng W et Song ZH, 2003)
	TM3	D3.49A	AM (Gi)	NB	
		R3.50A	AM (Gi)	IN	
		A6.34E	AM (Gi)	NB	
Cannabinoid receptor 1	TM3	L3.43A	CA	?	(D'Antona AM <i>et al</i> , 2006; D'Antona AM <i>et al</i> , 2006) Preuves indirectes
		T3.46I	CA	?	
		T3.46A	IN	----	
H1R	WT		CA	10	(Bakker RA <i>et al</i> , 2008) Activation de NF-κB
	TM6	I6.40R	CA	70	
		I6.40E	CA	80	
		I6.40G	CA	70	
		I6.40K	CA	100	
		I6.40F	CA	20	
		I6.40S	CA	90	
		I6.40A	CA	65	
GHSR		WT	AGI	17	(Liu G <i>et al</i> , 2007)
	TM3	I3.43T	AGI	83	
	TM4	V4.42M	DBA	17	
	TM6	F6.51L	DBA	34	
TSHR	TM1	S1.43I	CA	?	(Tao YX, 2008) Revue de littérature (Pas de résultats)
		G1.49S	CA	?	
	TM2	M2.43T	CA	?	
		M2.53V	CA	?	
	TM3	S3.36N,R	CA	?	
		V3.40A	CA	?	
		L3.43E,Q,R	CA	?	
	TM5	V5.54F,L	CA	?	
		Y5.58N	CA	?	
	TM6	D6.30G	CA	?	
		A6.34I,S,V	CA	?	
		M6.37I	CA	?	
		L6.40F	CA	?	
		I6.41L	CA	?	
		F6.42C,I,L	CA	?	
		T6.42I (?)	CA	?	
		T6.43A	CA	?	
		D6.44A,E,H, Y	CA	?	
		P6.50S	CA	?	

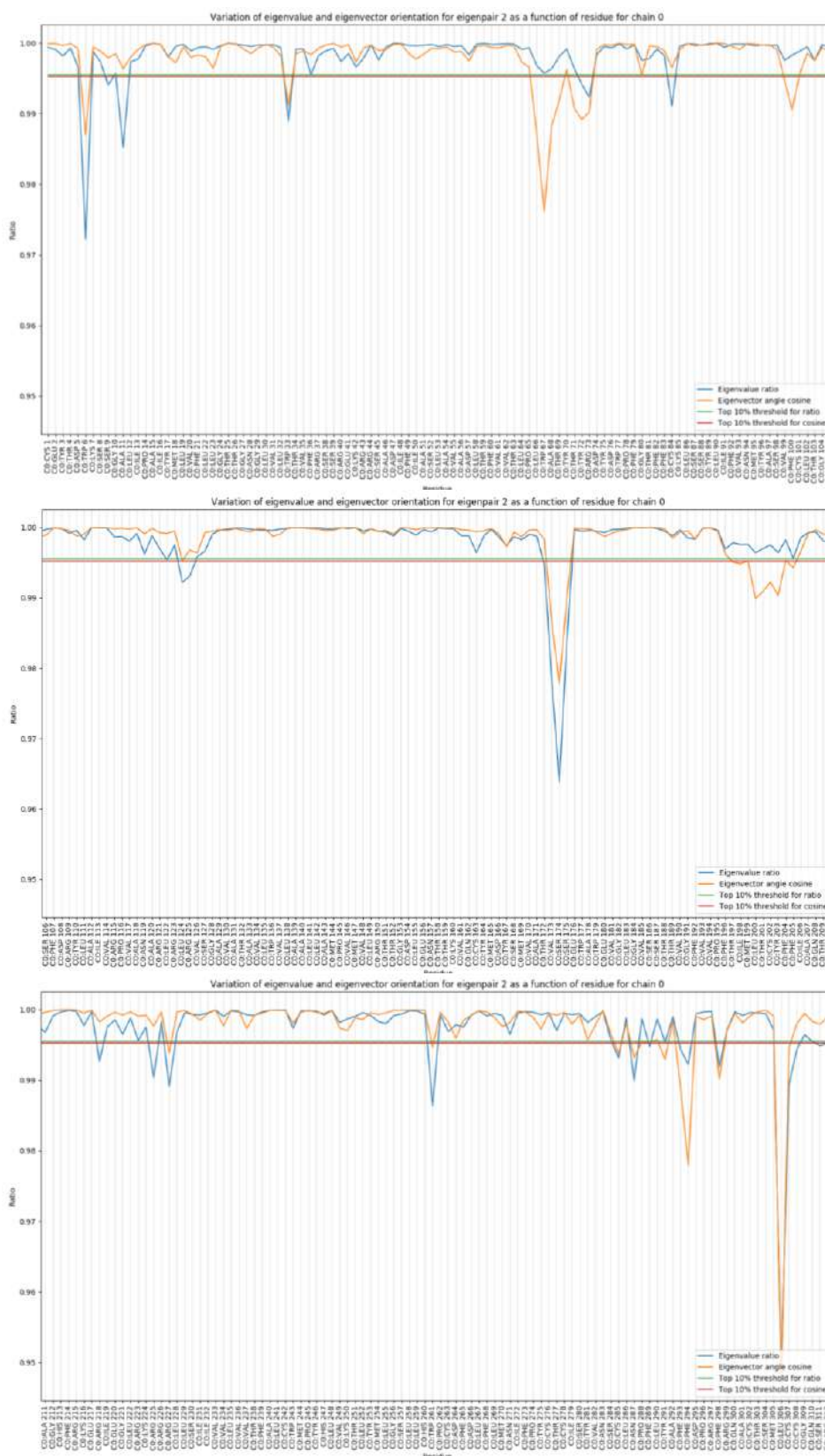
	TM7/H 8	F7.41L	CA	?	
		N7.45S	CA	?	
		C7.47Y	CA	?	
		L7.52V	CA	?	
LHCGR	TM1	L1.41P	CA	?	
		A1.46V	CA	?	
	TM2	M2.42T	CA	?	
	TM3	L3.43R	CA	?	
	TM5	I5.54L	CA	?	
	TM6	D6.30G	CA	?	
		A6.34V	CA	?	
		M6.37I	CA	?	
		A6.38V	CA	?	
		I6.41L	CA	?	
		T6.43I	CA	?	
		D6.44E,G,Y	CA	?	
		C6.45R	CA	?	
FSHR	TM3	T3.32A	CA	?	
	TM5	I5.45T	CA	?	
	TM6	D6.30G,N	CA	?	
MC4R	TM2	I2.62S,T	DBA	?	
	TM3	S3.30L	CA	?	
	TM5	F5.48L	DBA	?	
	TM6	N6.30S	DBA	?	
		L6.40Q	CA	?	
H4R	WT		AM (Gi)	100	(Schneider EH <i>et al</i> , 2010)
	TM3	R3.50A	AM (Gi)	0	
	TM6	A6.30E	AM (Gi)	100	
AT1	WT		pEC50 (Gq)	9.1	(Bonde MM <i>et al</i> , 2010) Biais à calculer avec le modèle équiactif, faute d'information sur le facteur de transduction et le facteur d'activité basale.
	TM3	D2.50N	pEC50 (Gq)	8.5	
		D3.49A/R3.50A	pEC50 (Gq)	Flat line	
	TM7	Y7.43F	pEC50 (Gq)	8.8	
		N7.49A	pEC50 (Gq)	8.6	
		Y7.53F	pEC50 (Gq)	8.8	
		WT	pEC50 (arr-3)	8.4	
	TM3	D2.50N	pEC50 (arr-3)	7.9	

		D3.49A/R3.50A	pEC50 (arr-3)	8.2	
	TM7	Y7.43F	pEC50 (arr-3)	7.9	
		N7.49A	pEC50 (arr-3)	7.7	
		Y7.53F	pEC50 (arr-3)	7.7	
		WT	AM (arr-3)	0.2	
	TM7	N7.49A	AM (arr-3)	0.17	
MC4R		WT	CA	100	(Wang ZQ, Tao YX, 2011)
	ICL1	H76R	CA	1750	
	TM3	D3.49N	CA	1500	
	TM1	I1.59R	IN/NB (α -MSH, β -MSH)		
	TM3	D3.49Y	IN/NB (α -MSH, β -MSH)		
	TM4	I4.45 Δ	IN/NB (α -MSH, β -MSH)		
	TM5	I5.40T	IN/NB (α -MSH, β -MSH)		
	TM6	P6.50Q	IN/NB (α -MSH, β -MSH)		
	TM7/H ₈	Q7.58X (?)	IN/NB (α -MSH, β -MSH)		

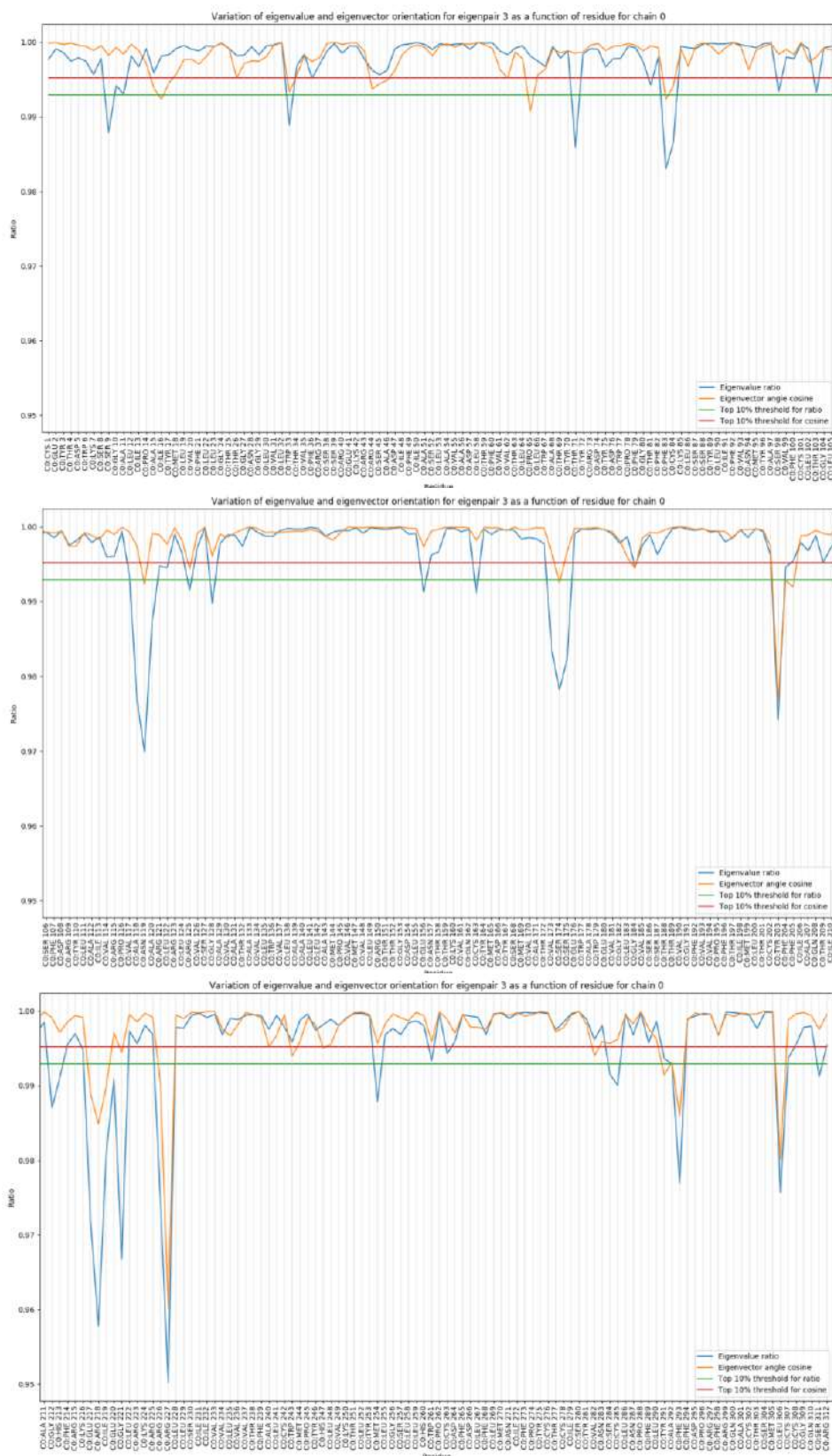
Annexe IV : Résultats de l'analyse par sous-matrices principales du système hAPJ-apo



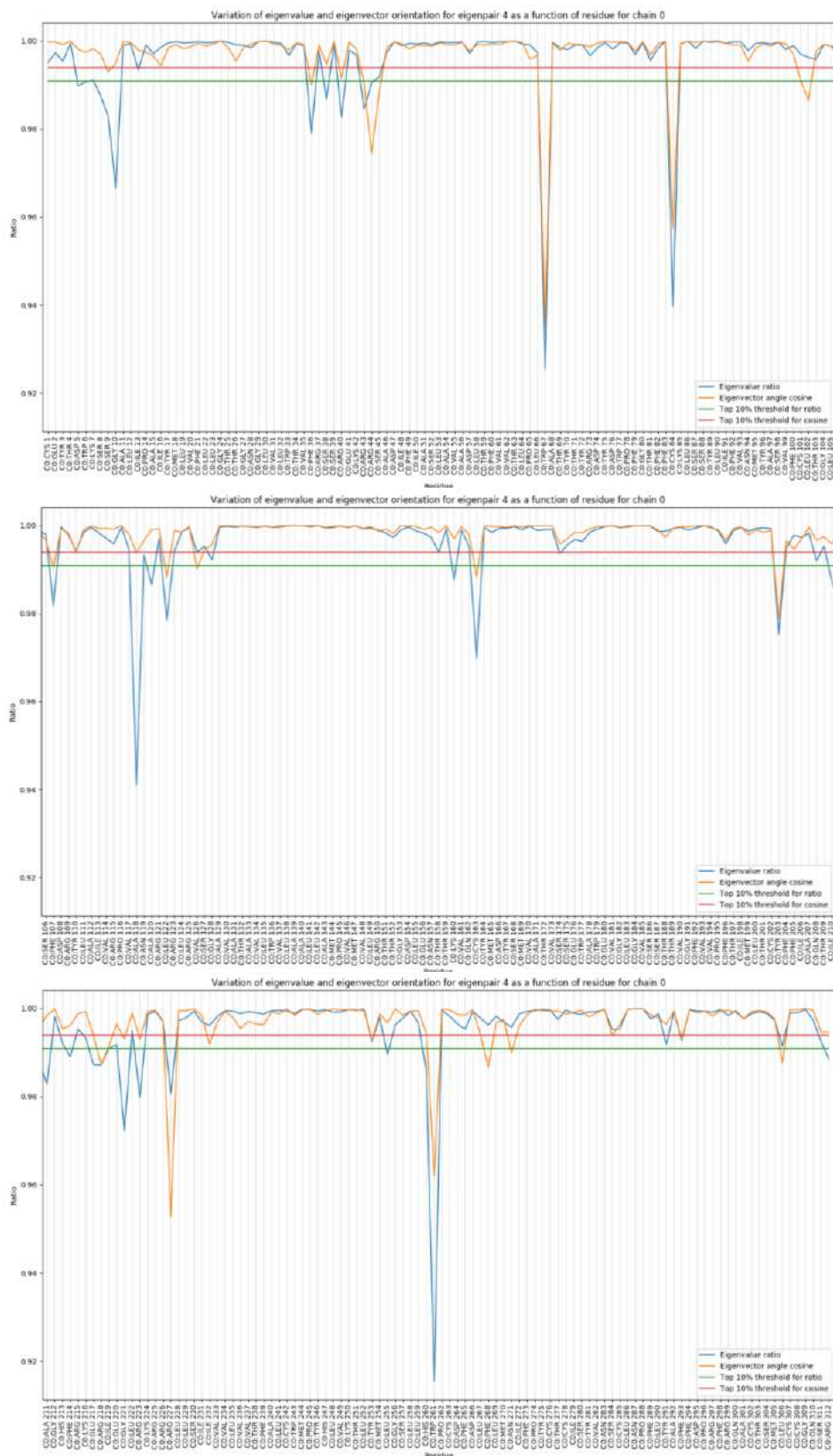
Résultat de l'analyse par sous-matrices principales de la composante indépendante 1 du système hAPJ-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



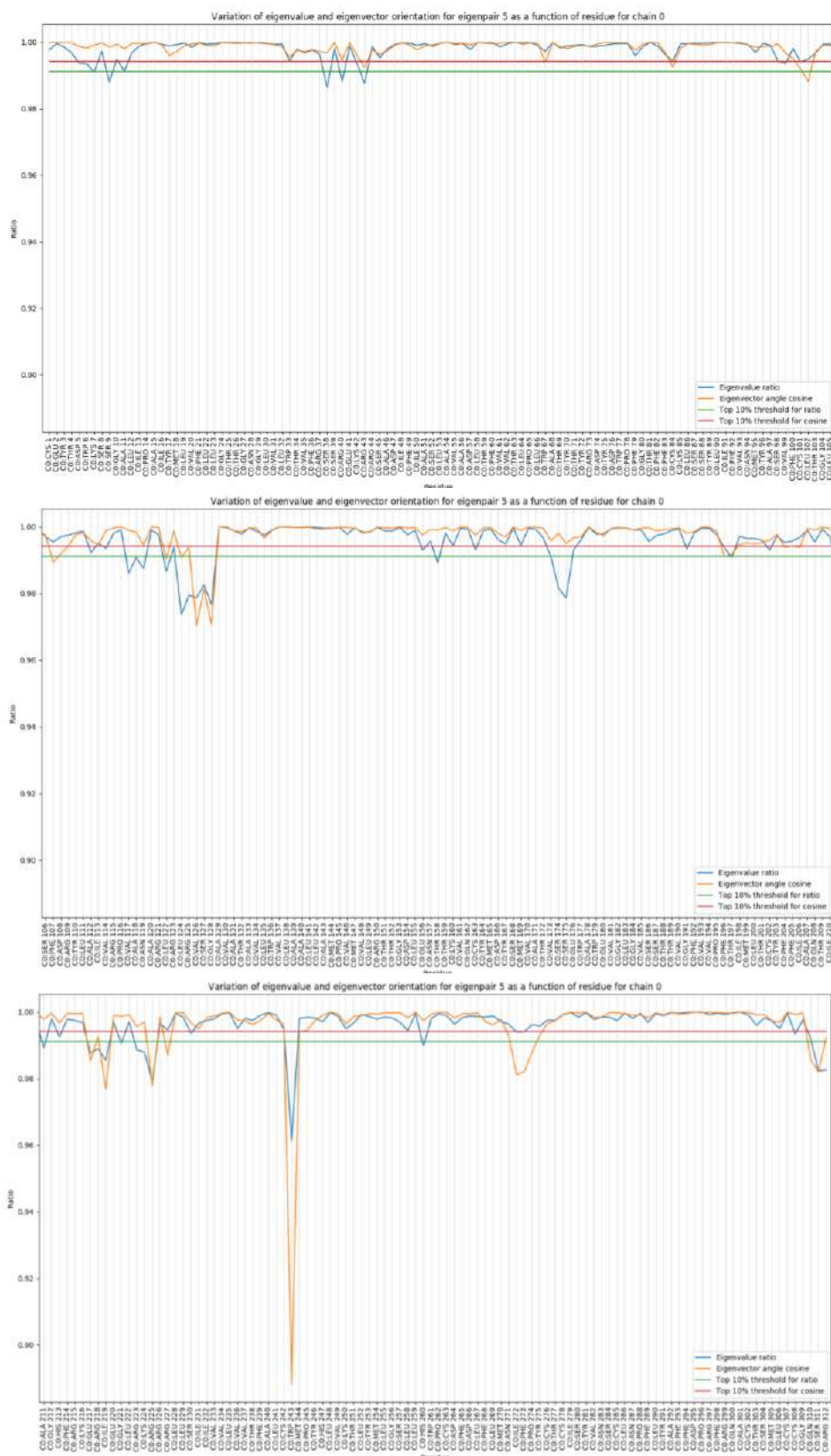
Résultat de l'analyse par sous-matrices principales de la composante indépendante 2 du système hAPJ-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



Résultat de l'analyse par sous-matrices principales de la composante indépendante 3 du système hAPJ-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

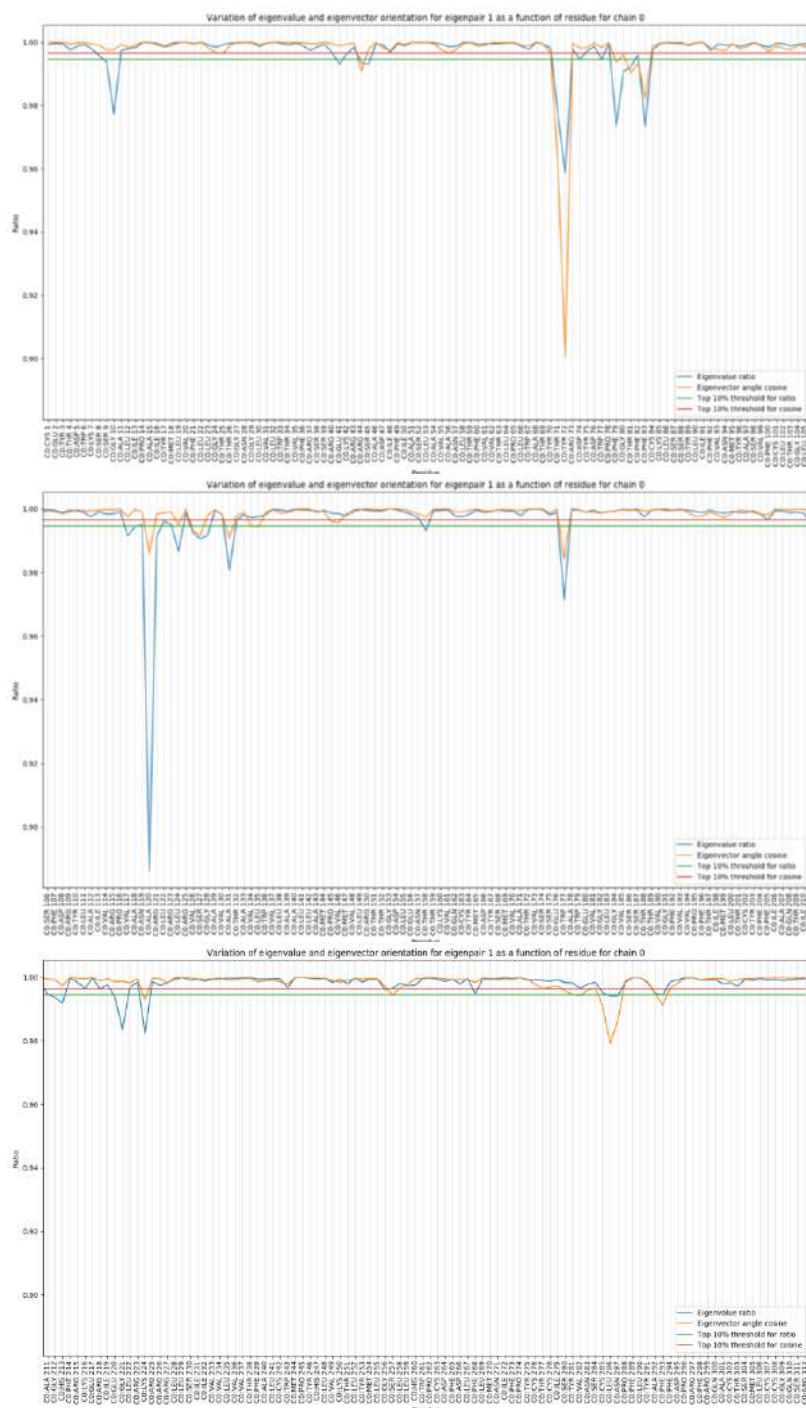


Résultat de l'analyse par sous-matrices principales de la composante indépendante 4 du système hAPJ-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



Résultat de l'analyse par sous-matrices principales de la composante indépendante 5 du système hAPJ-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

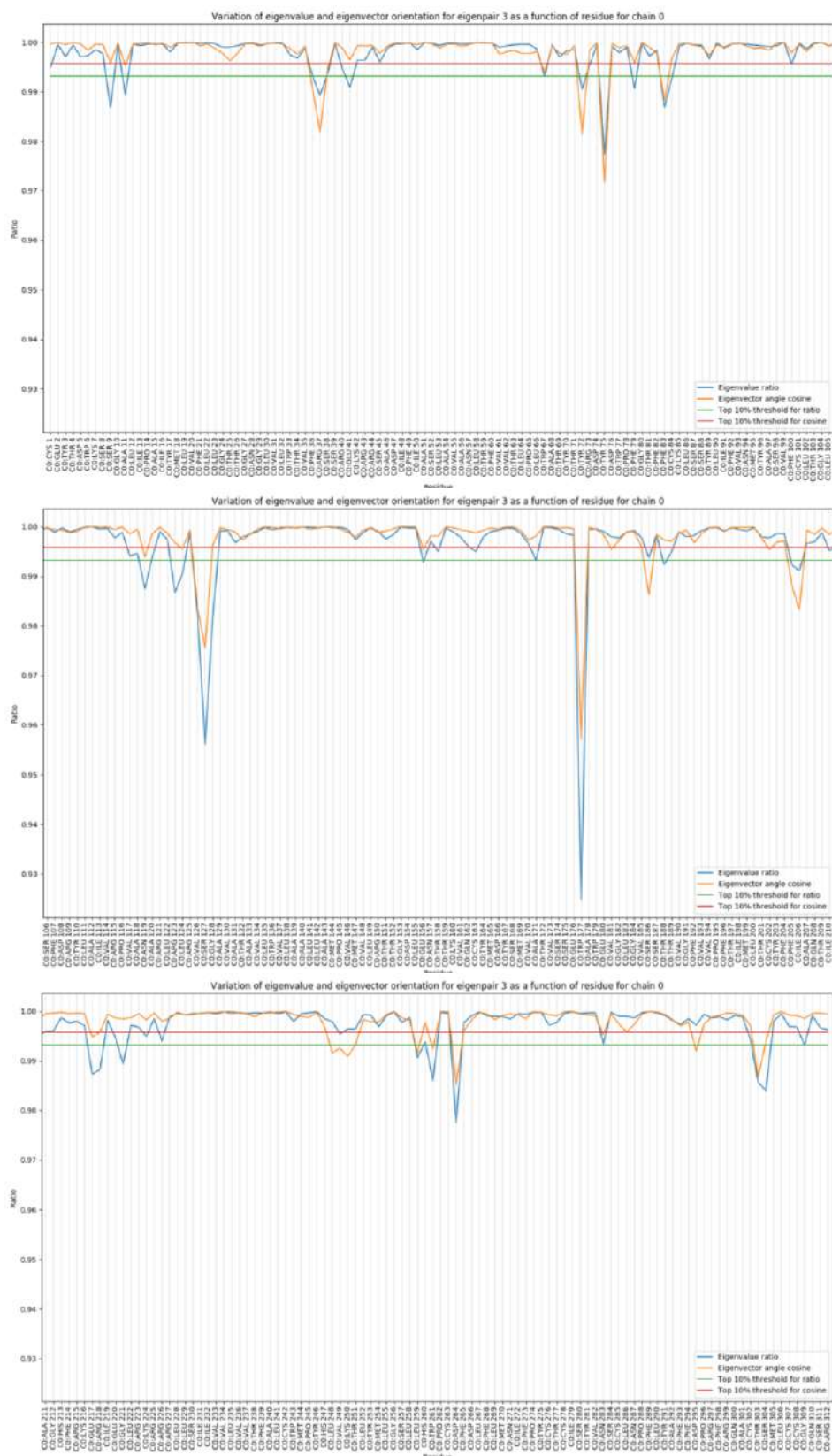
Annexe V : Résultats de l'analyse par sous-matrices principales du système hAPJ-D75N-apo



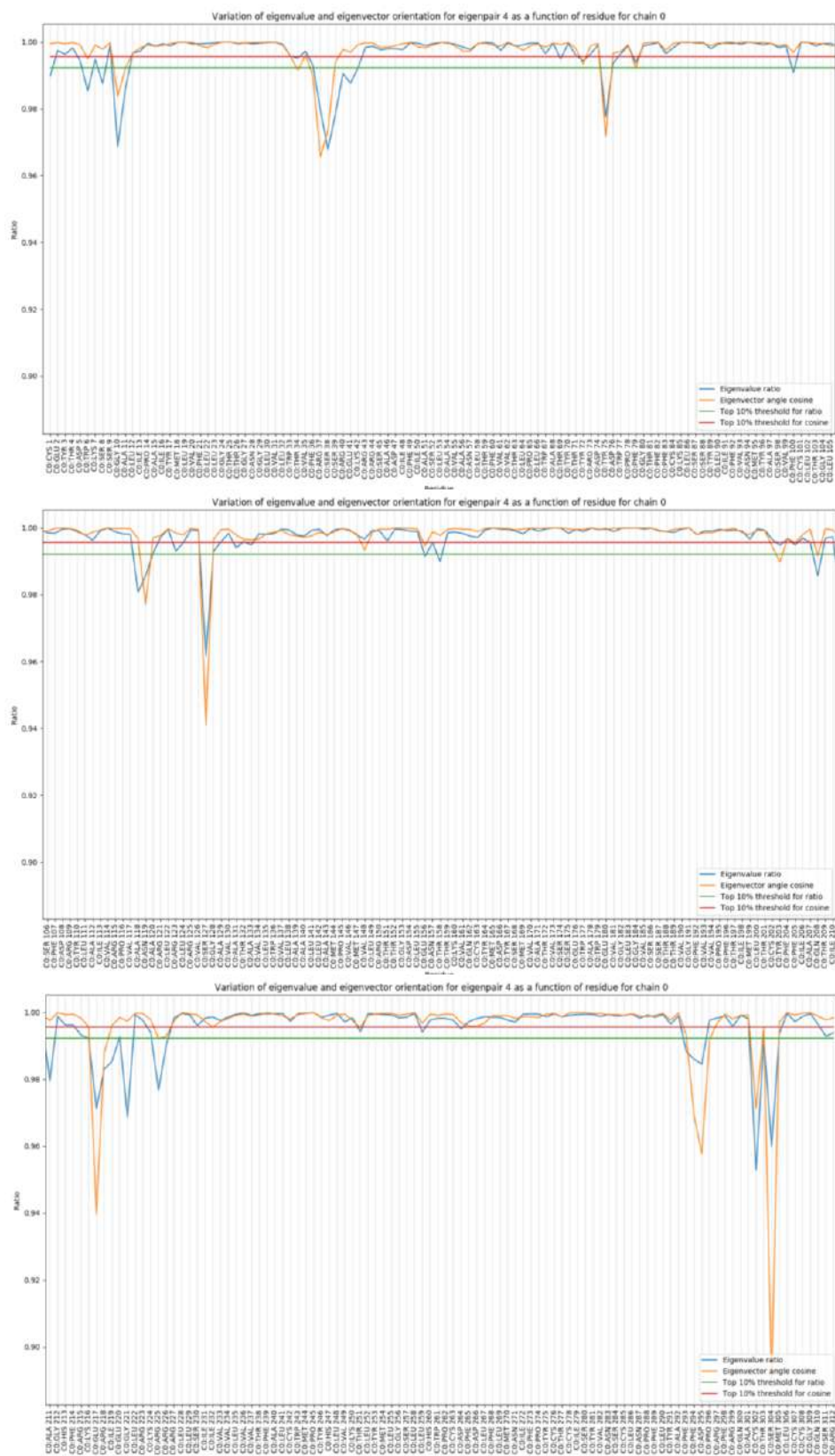
Résultat de l'analyse par sous-matrices principales de la composante indépendante 1 du système hAPJ-D75N-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



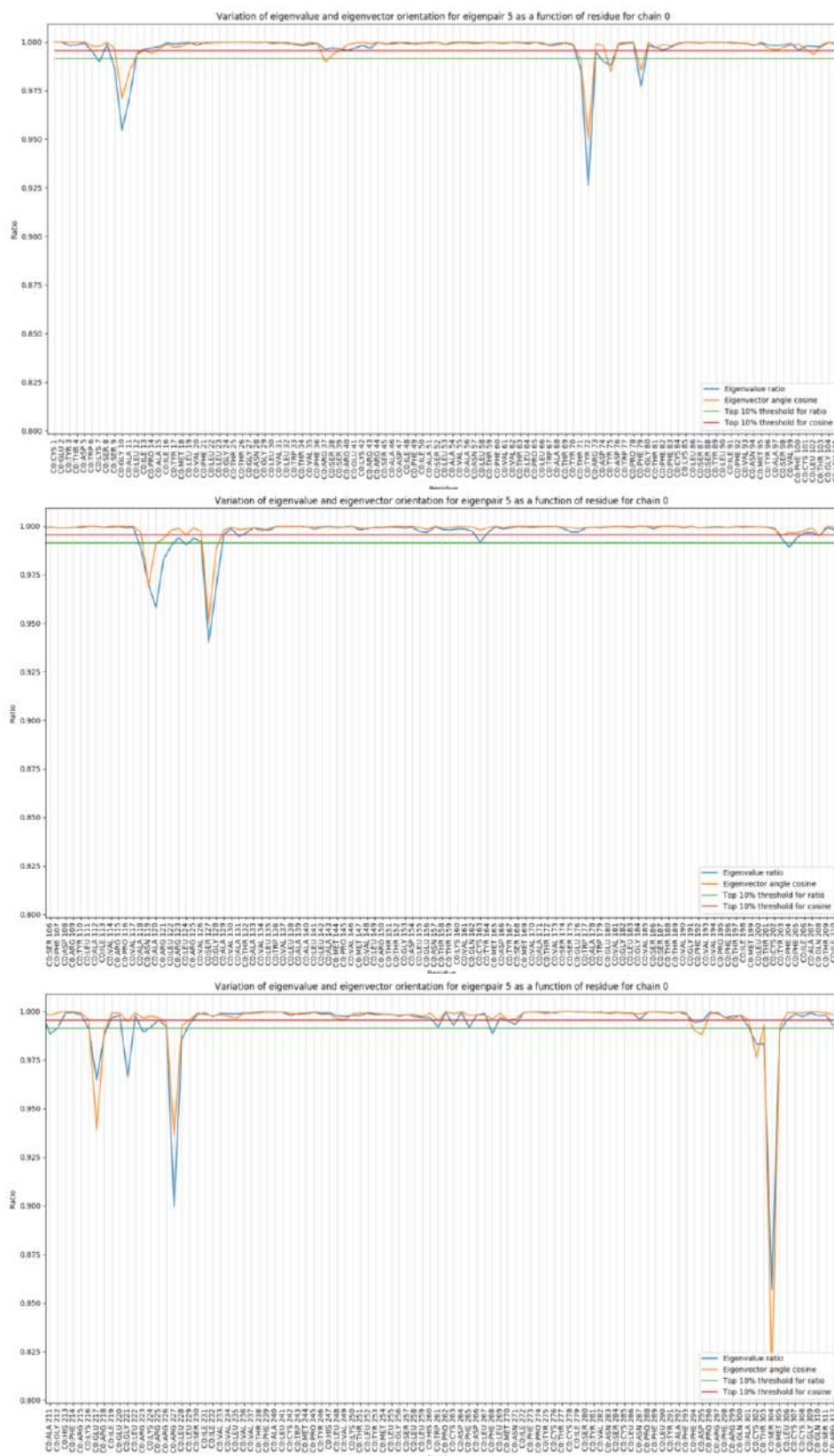
Résultat de l'analyse par sous-matrices principales de la composante indépendante 2 du système hAPJ-D75N-*apo*. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



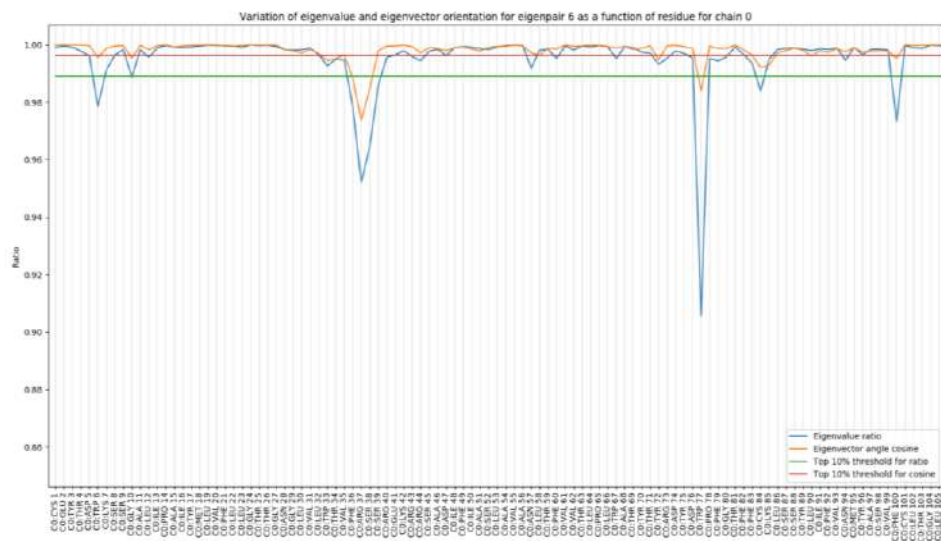
Résultat de l'analyse par sous-matrices principales de la composante indépendante 3 du système hAPJ-D75N- α . Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



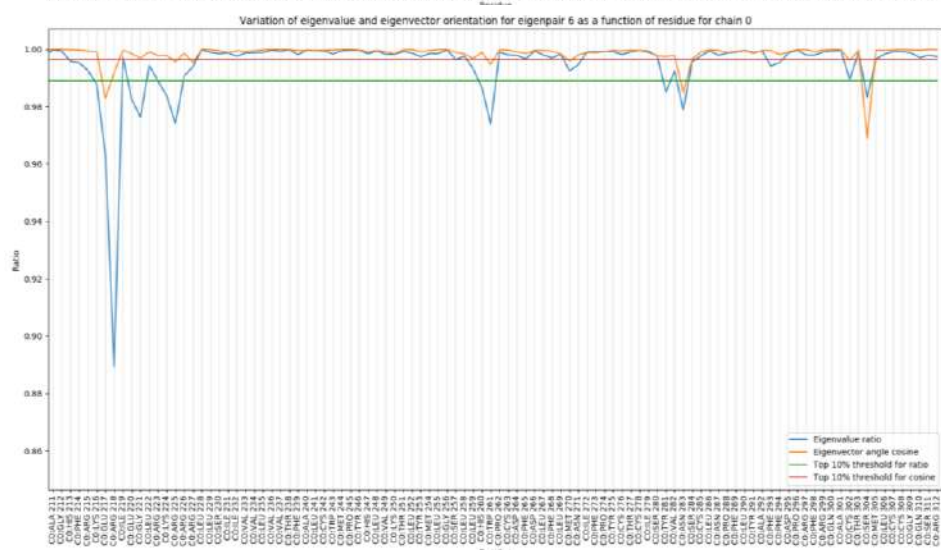
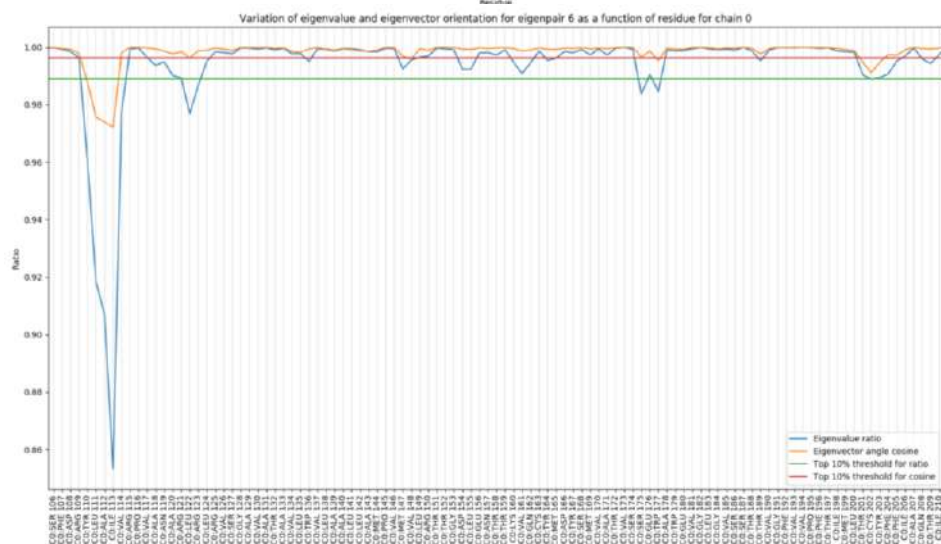
Résultat de l'analyse par sous-matrices principales de la composante indépendante 4 du système hAPJ-D75N-*apo*. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

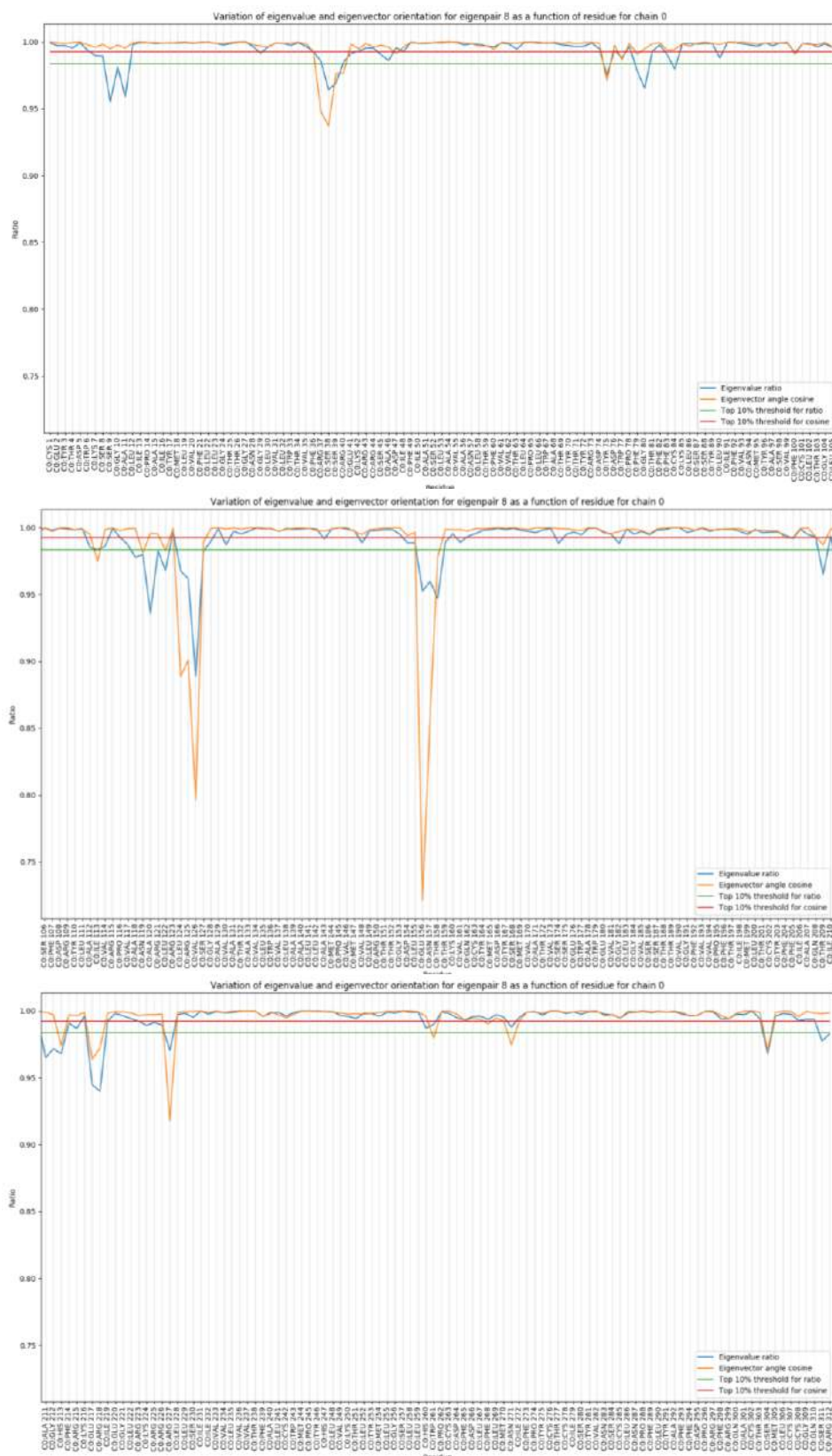


Résultat de l'analyse par sous-matrices principales de la composante indépendante 5 du système hAPJ-D75N- α . Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



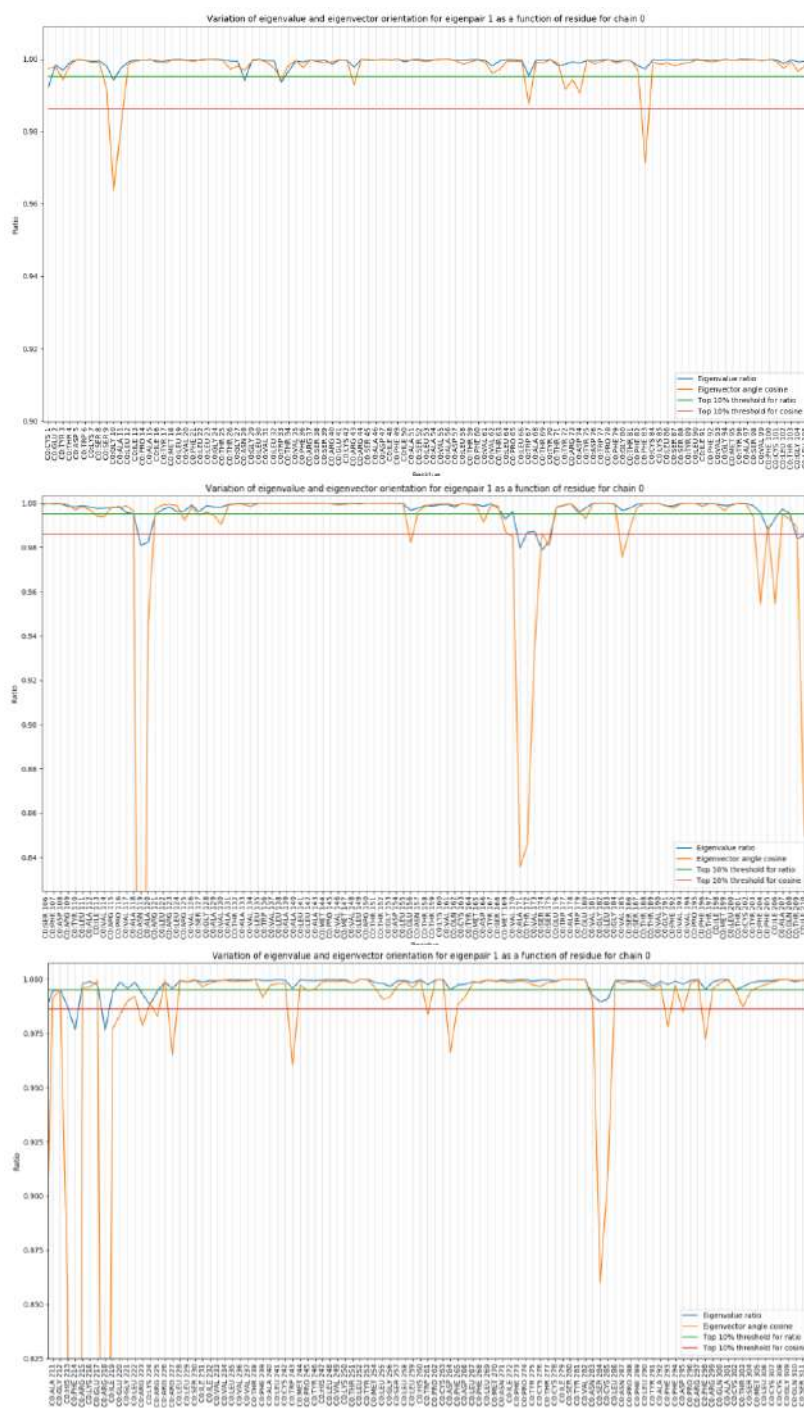
Résultat de l'analyse par sous-matrices principales de la composante 6 du système hAPJ-D75N-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



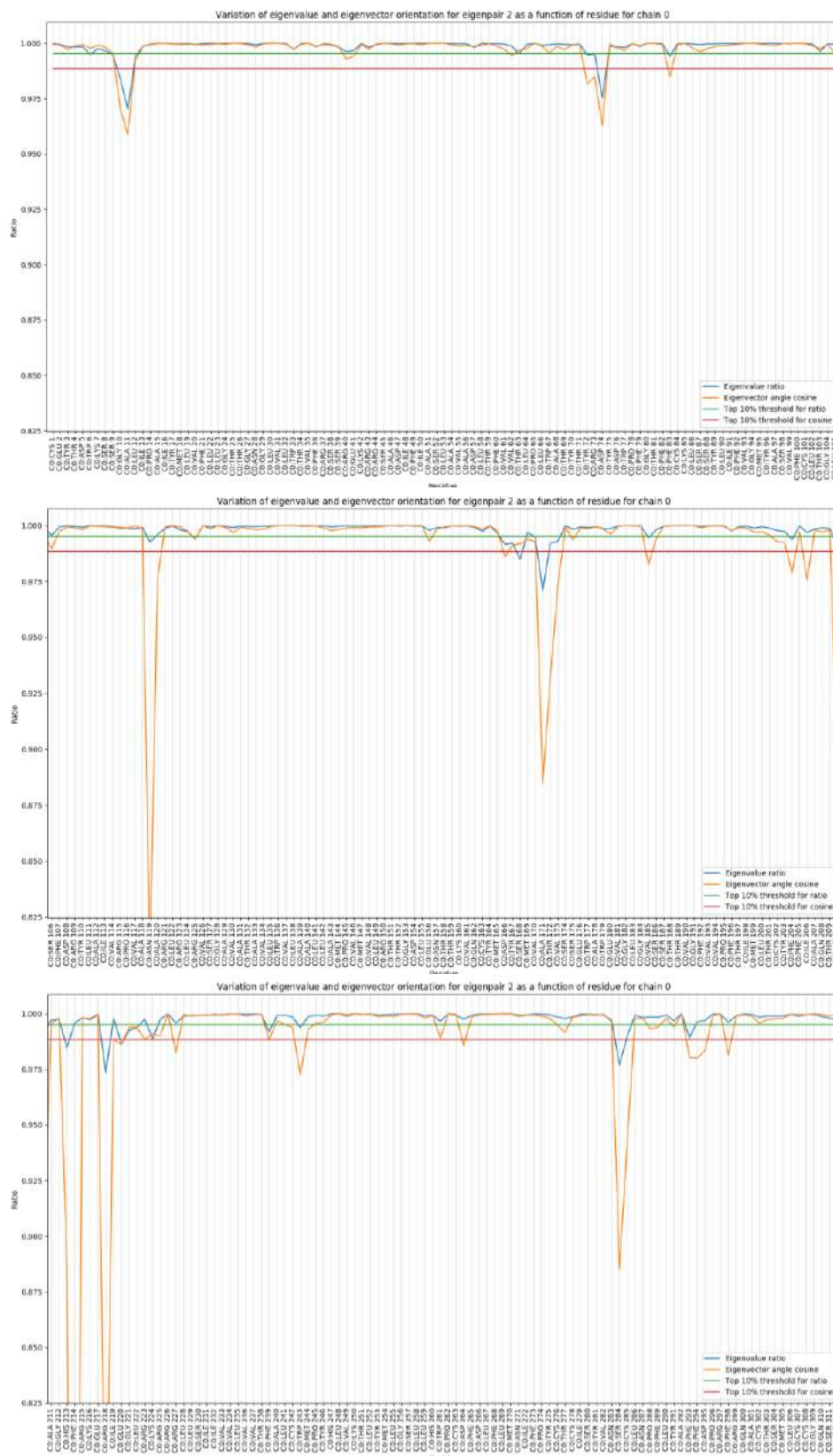


Résultat de l'analyse par sous-matrices principales de la composante indépendante 8 du système hAPJ-D75N-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

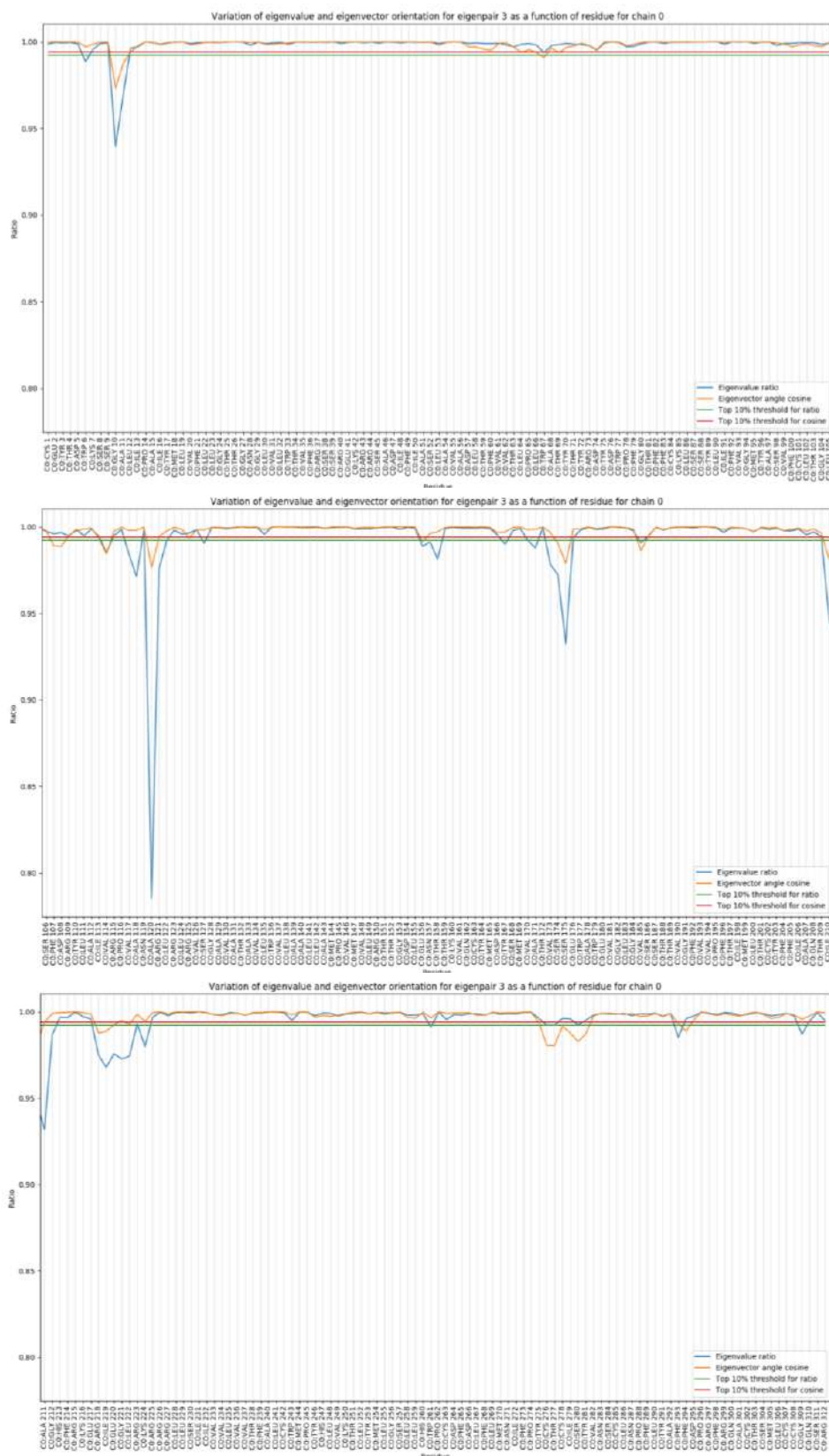
Annexe VI : Résultats de l'analyse par sous-matrices principales du système hAPJ-N112G-apo



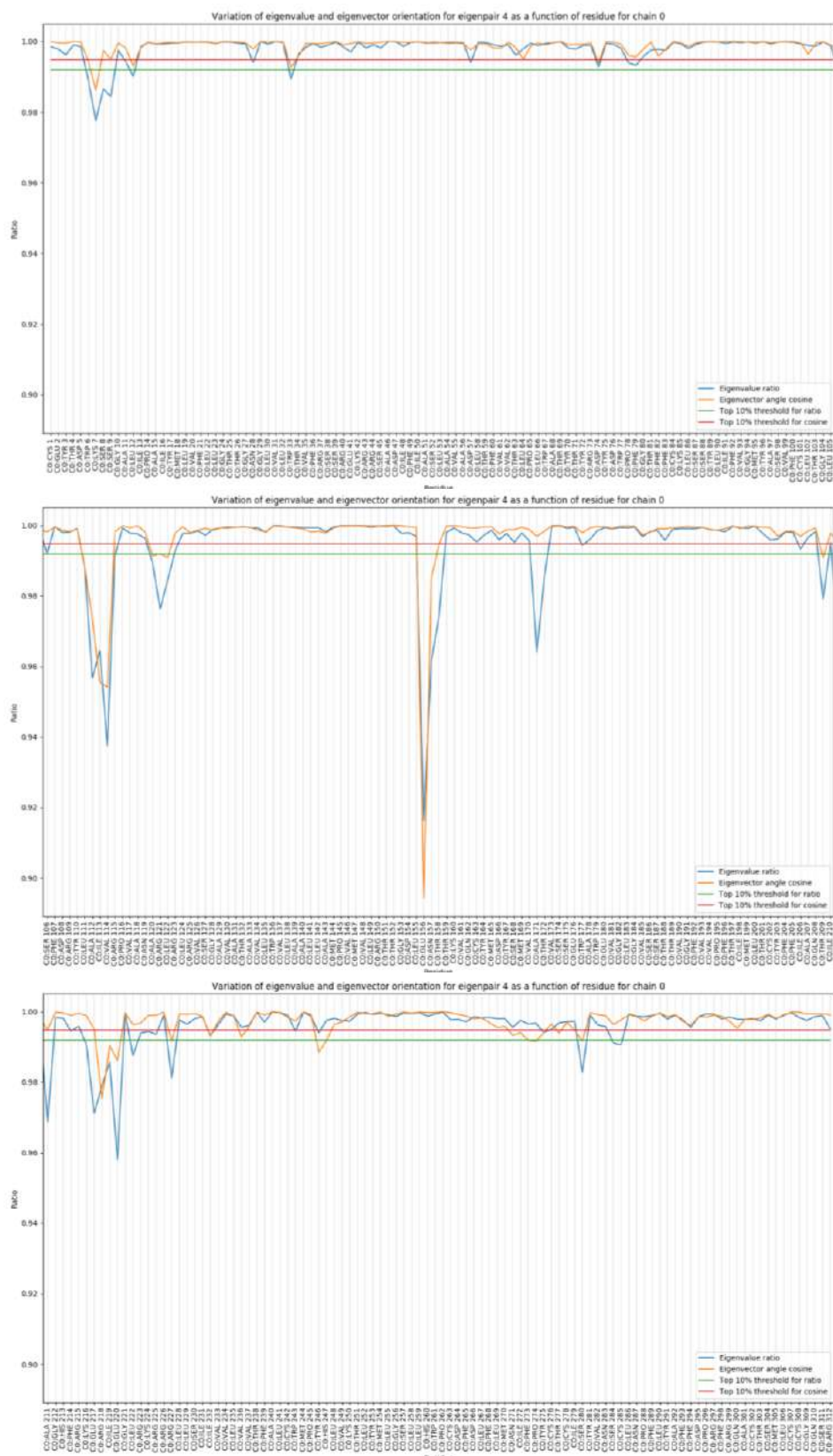
Résultat de l'analyse par sous-matrices principales de la composante indépendante 1 du système hAPJ-N112G-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



Résultat de l'analyse par sous-matrices principales de la composante indépendante 2 du système hAPJ-N112G-*apo*. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

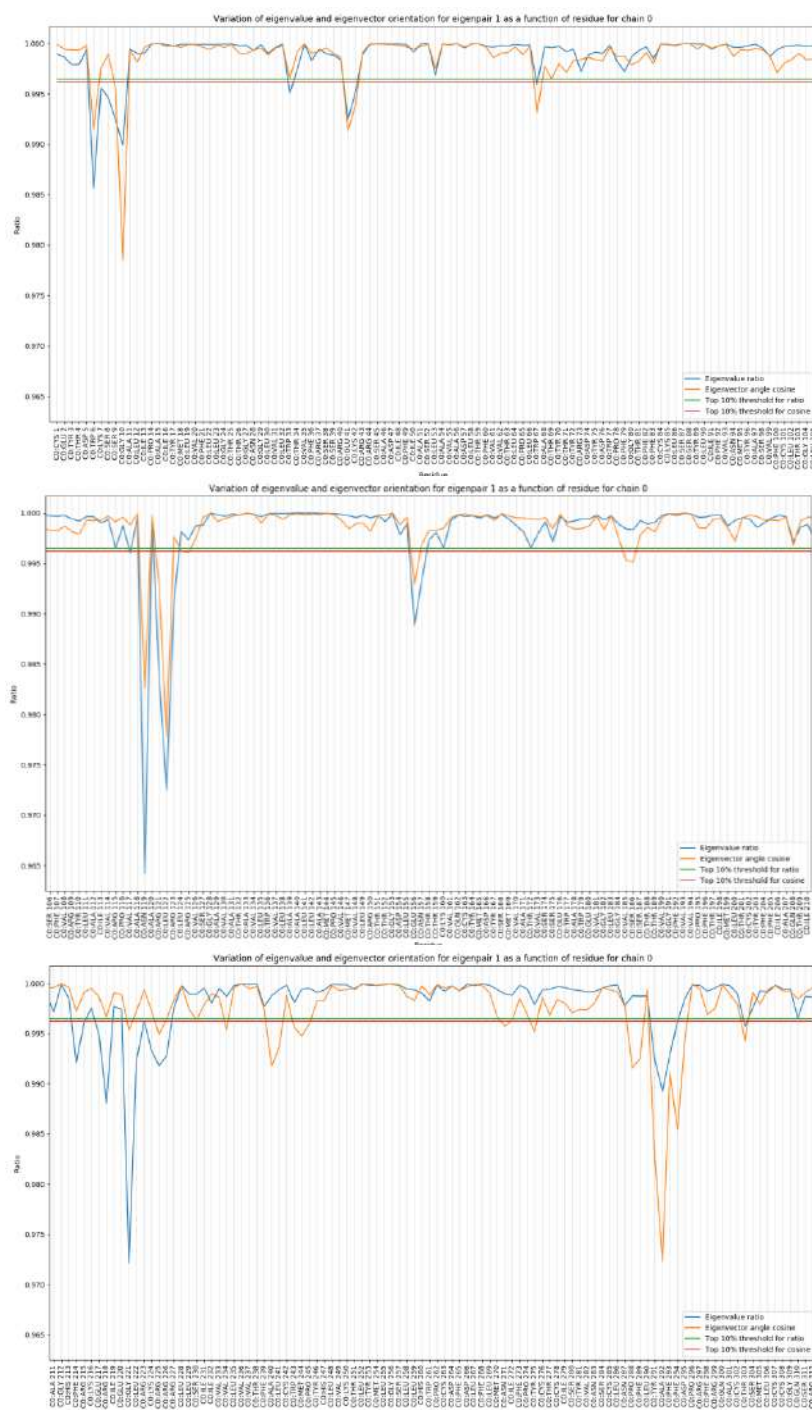


Résultat de l'analyse par sous-matrices principales de la composante indépendante 3 du système hAPJ-N112G-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

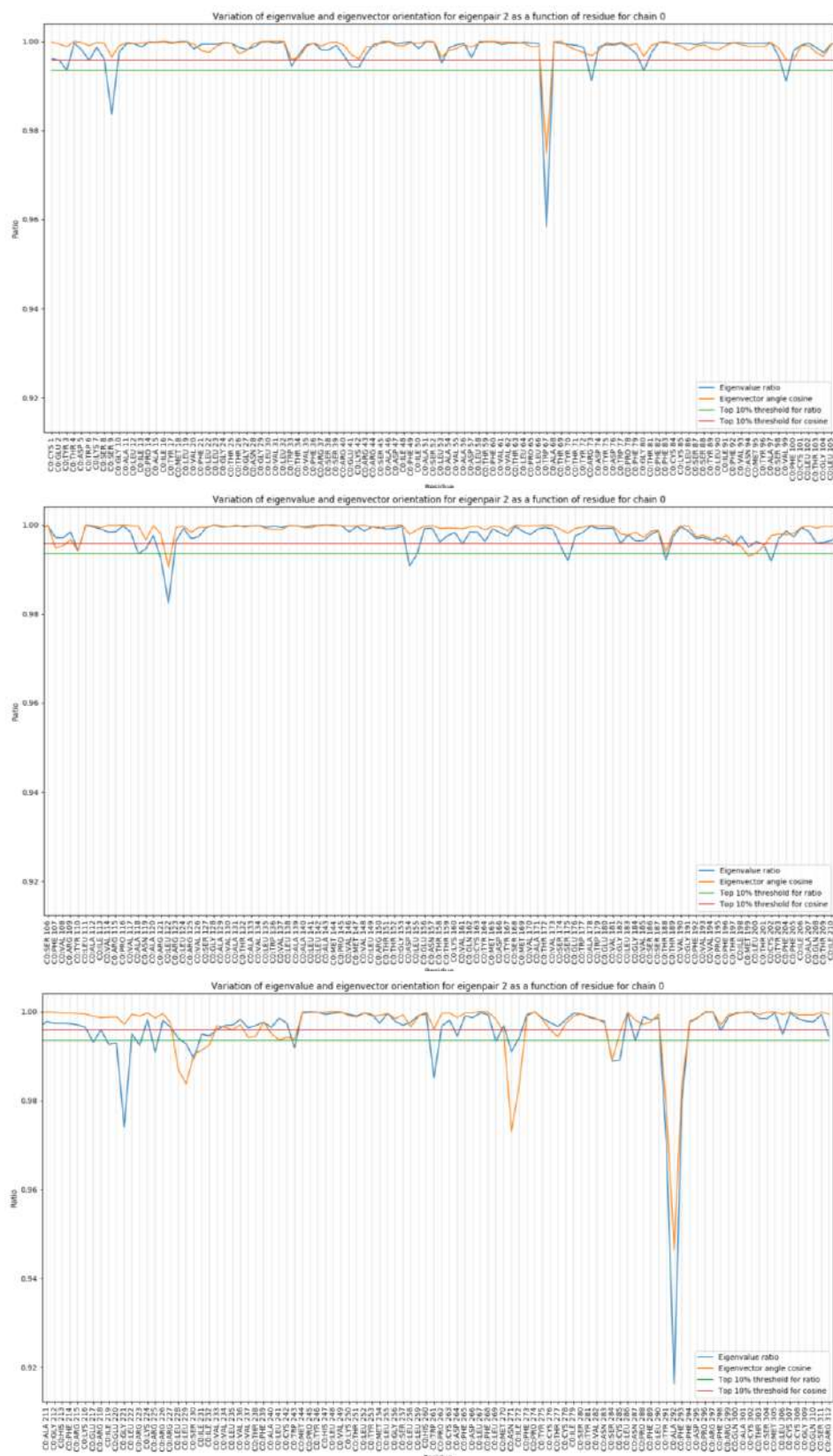


Résultat de l'analyse par sous-matrices principales de la composante indépendante 4 du système hAPJ-N112G-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

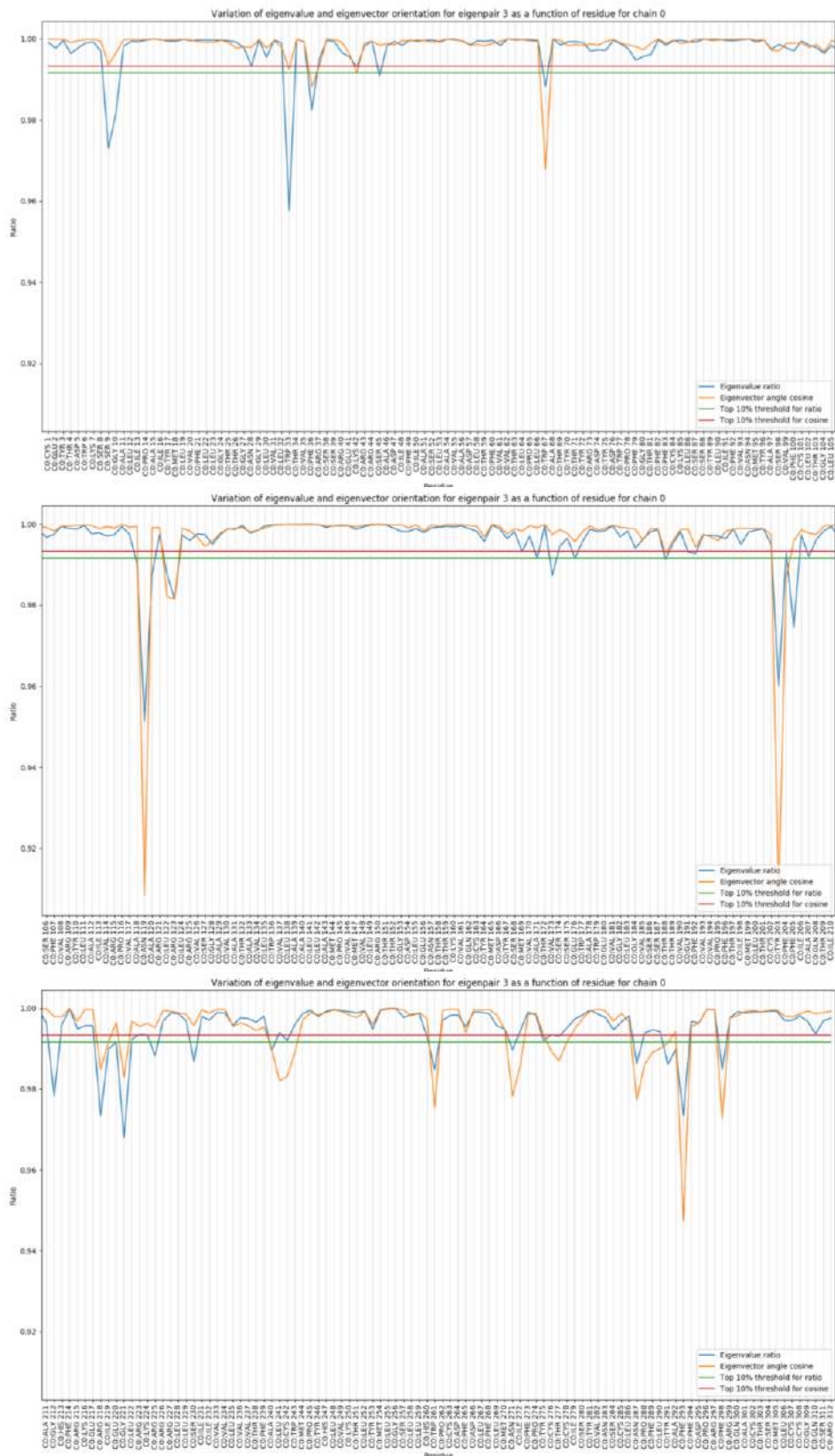
Annexe VII : Résultats de l'analyse par sous-matrices principales du système hAPJ-D126V-apo



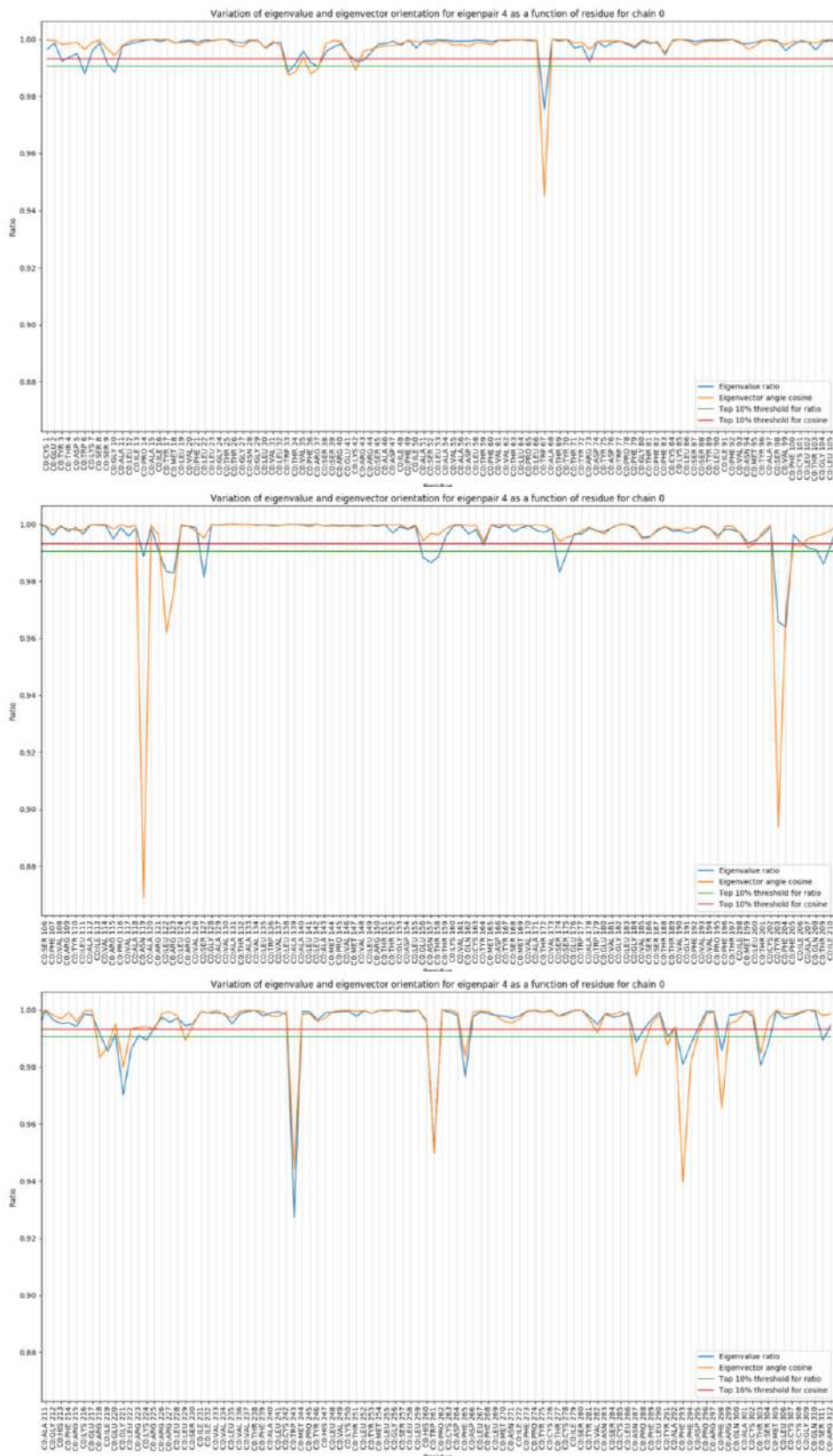
Résultat de l'analyse par sous-matrices principales de la composante indépendante 1 du système hAPJ-D126V-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



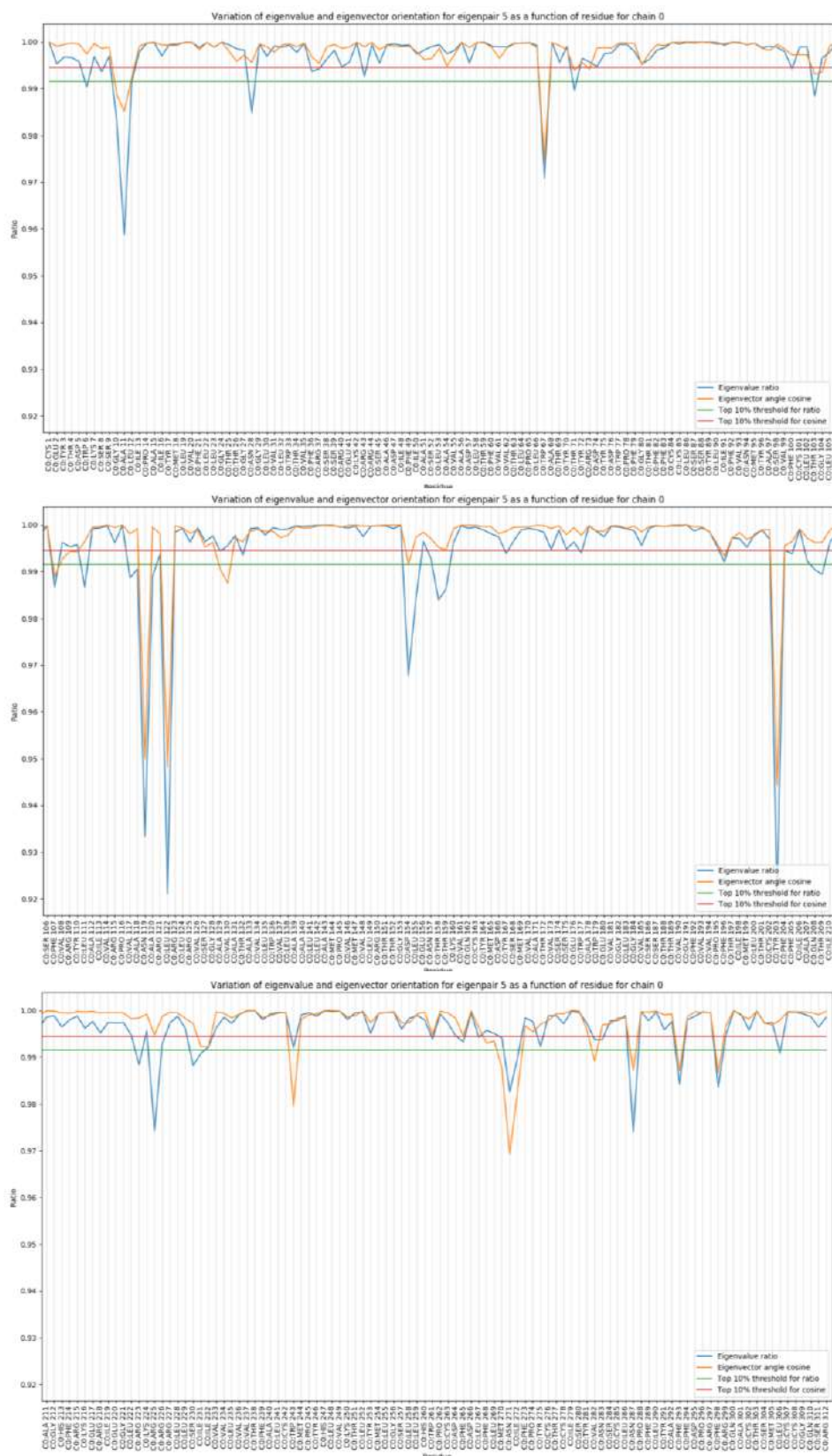
Résultat de l'analyse par sous-matrices principales de la composante indépendante 2 du système hAPJ-D126V-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



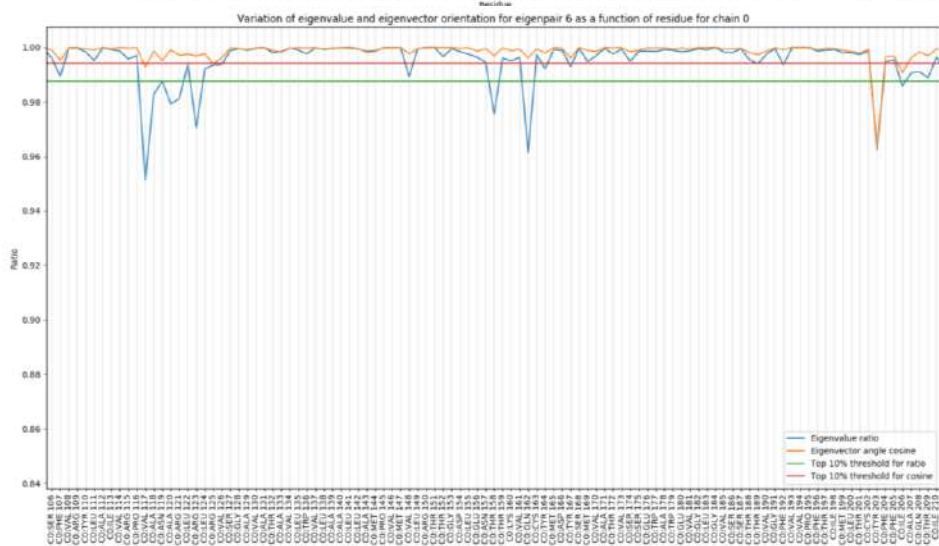
Résultat de l'analyse par sous-matrices principales de la composante indépendante 3 du système hAPJ-D126V- α . Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



Résultat de l'analyse par sous-matrices principales de la composante indépendante 4 du système hAPJ-D126V-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

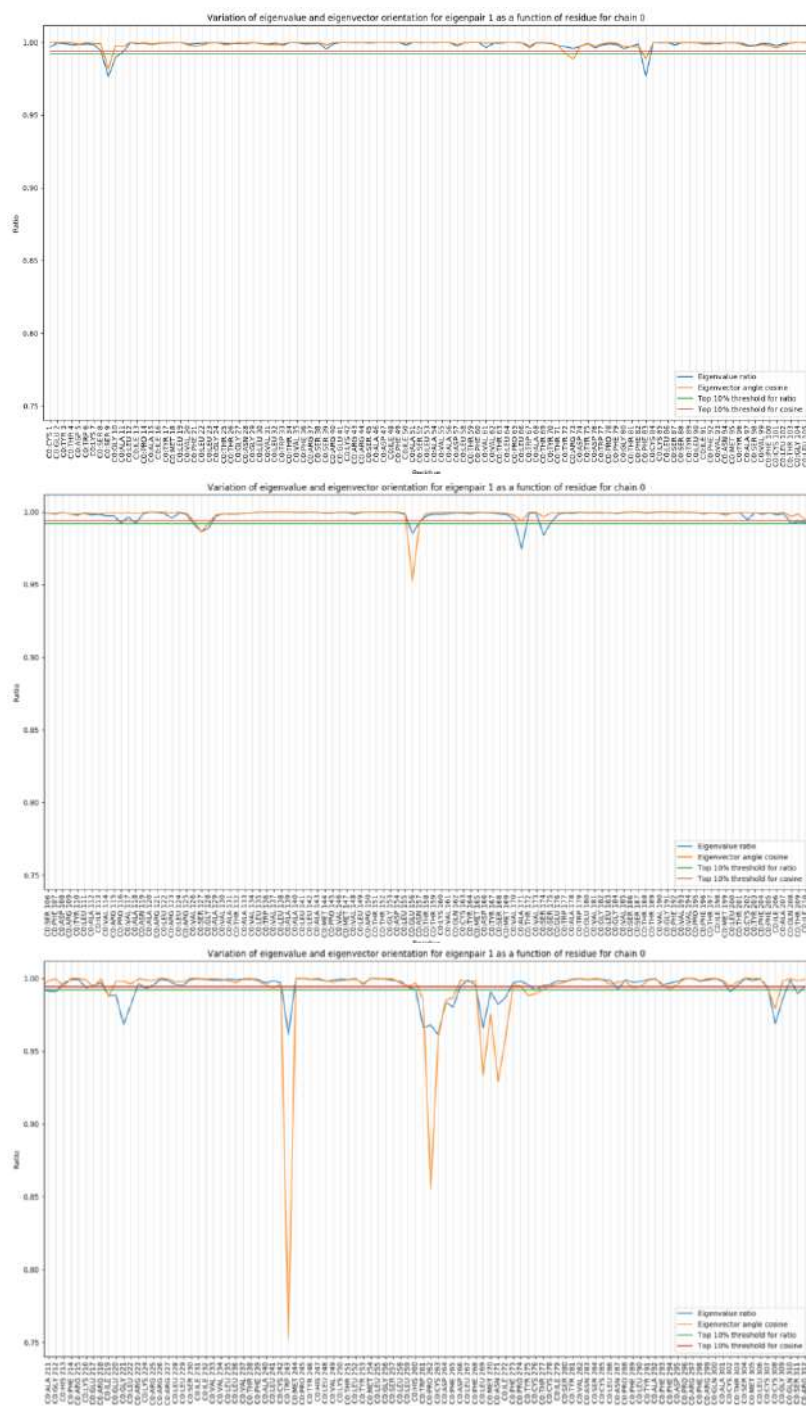


Résultat de l'analyse par sous-matrices principales de la composante indépendante 5 du système hAPJ-D126V-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

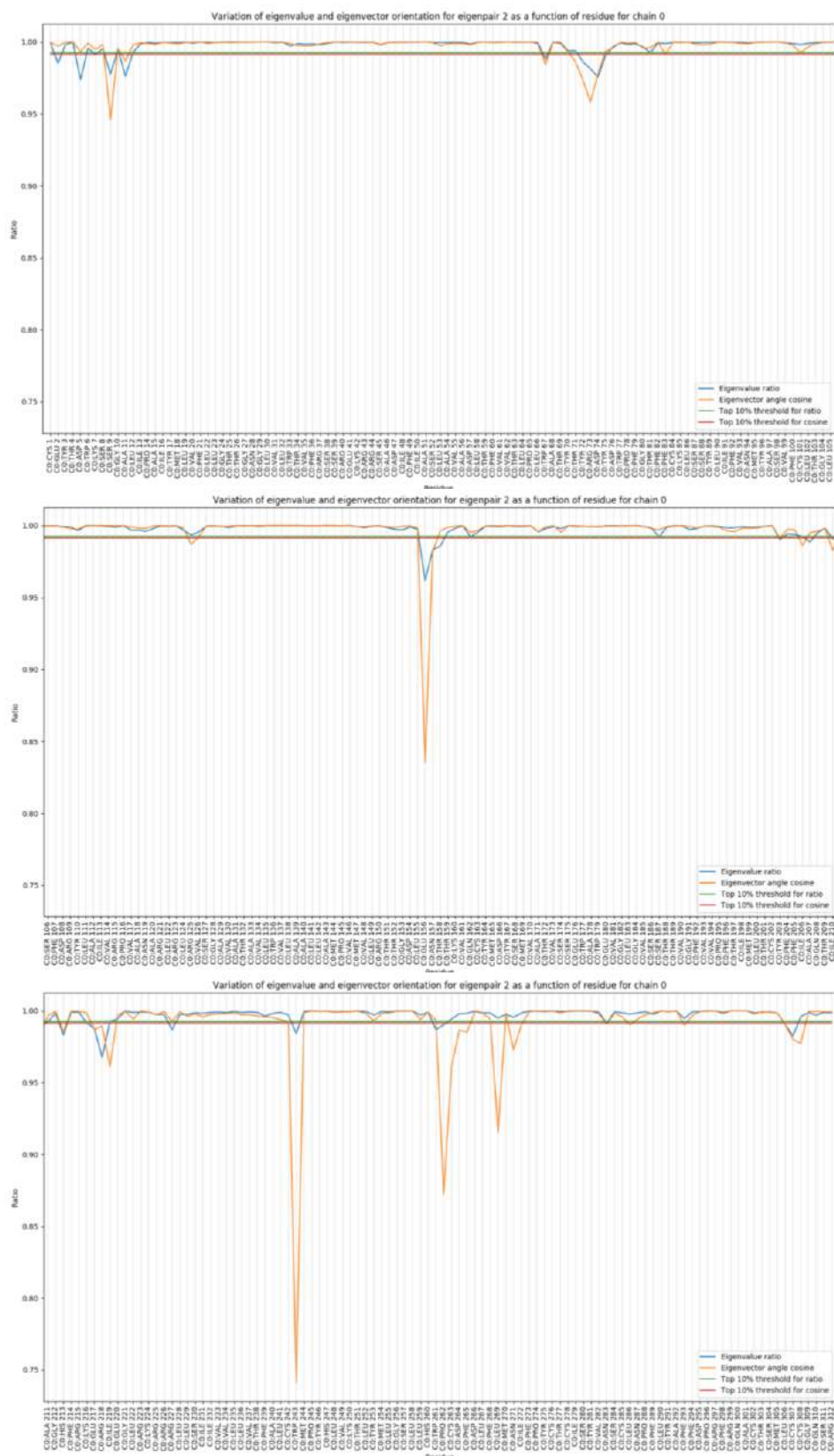


Résultat de l'analyse par sous-matrices principales de la composante 6 du système hAPJ-D126V-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.

Annexe VIII : Résultats de l'analyse par sous-matrices principales du système hAPJ-V254L-apo



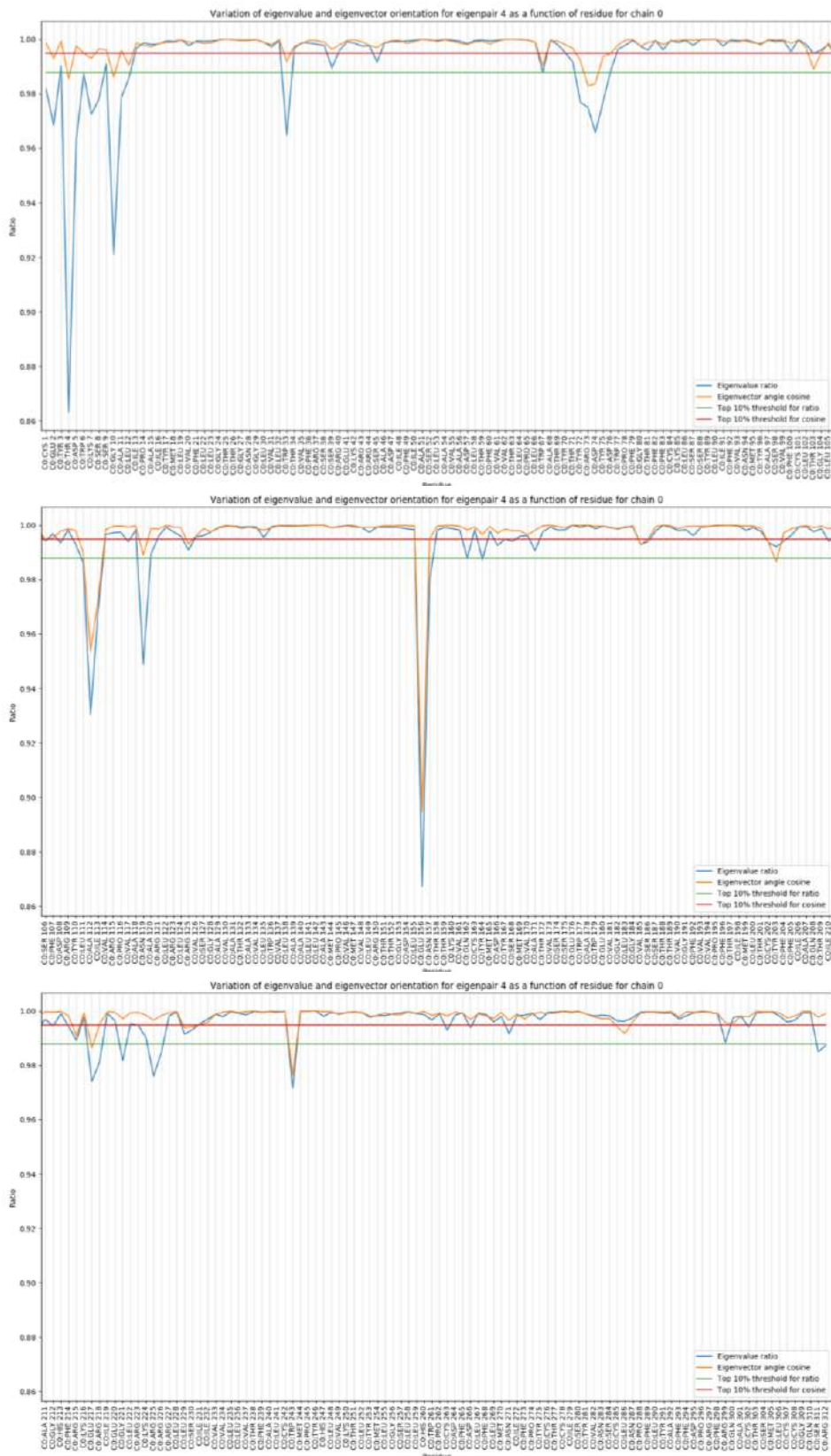
Résultat de l'analyse par sous-matrices principales de la composante indépendante 1 du système hAPJ-V254L-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



Résultat de l'analyse par sous-matrices principales de la composante indépendante 2 du système hAPJ-V254L-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



Résultat de l'analyse par sous-matrices principales de la composante indépendante 3 du système hAPJ-V254L- α . Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.



Résultat de l'analyse par sous-matrices principales de la composante 4 du système hAPJ-V254L-apo. Les numéros des résidus sont ici décalés de 18 car la portion N-terminale de hAPJ est ignorée.